2015

Detecting Well-being in Digital Communities: An Interdisciplinary Engineering Approach for its Indicators

Margeret A. Hall
University of Nebraska at Omaha, mahall@unomaha.edu

Detecting Well-being in Digital Communities:

An Interdisciplinary Engineering Approach for its Indicators

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für
Wirtschaftswissenschaften
am Karlsruher Institut für Technologie (KIT)

eingereichte

DISSERTATION

von
Margeret Hall, MA

Tag der mündlichen Prüfung:
Referent:
Korreferent:
Prüfer:

2015 Karlsruhe

Abstract

Progressive institutions are those which include the strategic interests of their constituents along with their own. While the interest to expand and develop metrics in this area has been expressed, the application thereof is constrained. One reason for this is the lack of appropriate indicators. Well-being, or the experience of feeling good and functioning effectively, is well-positioned to become this indicator. Highly granular traces of well-being can be extracted from digital footprints left in online social media. Given the predominance of the online self in the Internet age, such data is abundant and manifold. Before well-being can be applied several challenges need to be addressed. In particular, this includes the operationalizing of well-being measurements, the creation of a suitable implementation framework, the identification and refinement of suitable data, and the technical application of a platform for the implementation of such a system.

In this thesis, the challenges of defining, refining, and applying well-being as a progressive management indicator are addressed. The thesis approaches these challenges from a service logic perspective, namely transformative service research. The first part defines well-being and shows the usefulness of integrating well-being into the service value chain. The second part of the thesis concentrates on case studies applying information-driven well-being assessments to online social media data. The thesis advocates an unobtrusive data extraction and evaluation model entitled the Social Observatory. With a Social Observatory, it becomes possible to view highly granular, very personalized data left in digital traces by online social media users. For highly frequent and low-cost assessments of well-being, text analytics and sentiment analysis are proposed and evaluated in this context. The thesis shows that sentiment analysis provides reliable well-being data with low research(er) bias that can be viewed from many granularity levels. A subsequent finding in this thesis is that is it possible to mitigate the bias introduced by individuals in their online profiles by isolating aspects of the users' personality.

The final part of this thesis holistically investigates a university's online social media network for its digital traces of communal well-being. The corresponding case study established that communal well-being can be detected and isolated as an indicator. Well-being, whilst generally existing as a baseline, is observed having spikes and dips that are directly related to events and incidents impacting the campus community. In particular, the concept of communal belongingness is a representative proxy of communal well-being; its longitudinal observation can be implemented as a tool of progressive community management.

This work's implications and contributions are highly relevant for service research as it advances the integration of consumer well-being and the service value chain. It also provides a substantial contribution to policy and strategic management by integrating constituents' values and experiences with recommendations for progressive community management.

Table of Contents

List of abbreviations

AMT	Amazon Mechanical Turk
API	Application Programming Interface
CDU	Christian Democratic Union
CSU	Christian Social Union
CMB	Common Method Bias
df	Degrees of Freedom
DRM	Daily Reconstruction Method
FDP	Free Democratic Party
GDP	Gross Domestic Product
GLM	Generalized Linear Model
GSS	General Social Survey
HCI	Human-Computer Interaction
HF	Human Flourishing
HFS	Human Flourishing Score
HIT	Human Intelligence Task
ISSP	International Social Survey Programme
JSON	JavaScript Object Notation
KiB	Kibibyte
KIT	Karlsruhe Institute of Technology
KMO	Kaiser-Meyer-Olkin
LIWC	Linguistic Inquiry and Word Count
MiB	Mebibyte
OECD	Organisation for Economic Cooperation and Development
PCA	Principal Component Analysis
PWB	Psychological Well-being
RMSE	Root-mean-square-error
RQ	Research Question
SD	SD
SDP	Social Democratic Party
SDT	Self-determination Theory
SQL	Structured Query Language
SWB	Subjective Well-being
TSR	Transformative Service Research

List of figures

List of tables

List of symbols

w_i	Weighting Factor
γ^{t-j}	Discounting Factor
CR_j	Certain Reward at Time Point *j*
EV_j	Expected Value of an Action at Time Point *j*
RPE_j	Difference Between Expected and Actual Rewards at Time Point *j*
U(Y,t)	Utility Estimate of Self-reported Well-being
H[U(Y,t)]	Continuous Non-differential Function Estimating Difference Between Actual and Self-reported Well-being
ϵ	Error Term
pe	Positive Emotion
I_c	Index of Positive Characteristics
I_f	Index of Positive Functioning
c_j	Positive Characteristics Construct
f_k	Positive Functioning Construct
P_c	Degree of Positive Characteristics
P_f	Degree of Positive Functioning
m	Item Traits of Positive Characteristics
l	Items Traits of Positive Functioning
$\mathrm{d}(\mathrm{x}, \mathrm{y})$	Distance Function Between the Variables *x* and *y*
ρ	Spearman's rho
α	Estimate of the Dispersion Parameter
β_0	Intercept of the Linear Regression
β	Regression Coefficient
$\mathrm{p}(\mathrm{x}; \mathrm{b}, \mathrm{w})$	Conditional Mean
n, x, y	A Numeric Variable
k	A Number in a Feature Space

Part I.
Introduction

Chapter I Introduction

"Happiness is based on a just discrimination of what is necessary, what is neither necessary nor destructive, and what is destructive."

The Ones Who Walk Away from Omelas (Le Gain, 1973)

1.1 Motivation: Well-being in Institutional Management

Without a doubt, the characteristics of the modern economy are that services are more foundational than ever (*servicitization*), modern institutions are more and more concerned with (human) factors outside of profitability (*humanization*), and that the Internet has become the kingmaker of it all (*digitalization*). The internet has enabled service providers to migrate and proliferate online as barriers to market entrance are significantly lowered (*OECD* 2010). It has also increased the stakes of institutional reputation maintenance by increasing transparency and participation, where institution is broadly defined as any persistent structure(s) that govern behavior (e.g., governments, social networks, companies) (Auer 2011; Friedman, Kahn Jr., and Borning 2003; Friedman 1996), (and is used synonymously with community in this thesis). Anyone with a smart device or internet connection becomes an experiential expert. Online reputations in turn become a valuable tool to expand and protect existing consumer[1] bases (Burke, Marlow, and Lento 2009). The touch of a button and a well-placed '#' can make or break a reputation, elect presidents, fund research for rare diseases, track (war) criminals, or even fell governments (Stieglitz and Dang-Xuan 2012; Skoric 2012; Böcking, Hall, and Schneider 2015). A consequence of this dynamic is a foundational reassessment by institutions of the means and ways of competition with respect constituent interactions. Increasing transparency and decreasing entry barriers necessitates that institutions not only properly service their constituents, but do well by them. The changeover of servicitization, humanization, digitalization can be enveloped by the term 'progressive community management' (Stiglitz, Sen, and Fitoussi 2009; Hall et al. 2012).

[1] Constituent, community member, and consumer are used interchangeably.

Implicit in these broad themes is that the relationship between institution and constituent is more personal than ever before. From this basis, the institution is able to assess not only traditional indicators like agency loss or turnover, but satisfaction, quality, and constituent emotional connectivity (Rosenbaum et al. 2011). The ability to foster and maintain direct relationships is oftentimes a direct consequence of the ease of information exchange and networking and lowered participation barriers afforded by digitalization (Vargo 2009; Rosenbaum et al. 2011; Dimitrova et al. 2011).

That what the World Bank calls development "beyond economic growth"[2] is increasing realization that human factors are considered a new norm in the assessments of institutional identity, policy, and overall health (Anderson et al. 2013; Norman and MacDonald 2004; Stiglitz, Sen, and Fitoussi 2009; Cameron, Bright, and Caza 2004). This is due in part to the fact that digitalization and digital tracks of relationships and interactions makes it easier for institutions to measure their impact on individuals. This has been positively influenced by digitalization. Institutions are finding it in their interests to monitor and respond holistically to indicators of both happiness and well-being of their stakeholders (Harter, Schmidt, and Keyes 2003). With the realization that the profit-first 'traditional bottom line' is no longer the final, nor the preferential goal of the modern economy (Norman and MacDonald 2004), institutions are incentivized to care about and invest in so-called human factors: social, ethical, and environmental reputations (Stiglitz, Sen, and Fitoussi 2009; Cameron, Bright, and Caza 2004). Far from the "race to the bottom" feared during the first years of globalization (Drezner 2004), digitalization of public spaces is instead a stable mechanism empowering individuals to document experienced positive and negative interactions served to them by institutions. The ubiquity of internet-enabled devices makes it increasingly easier to laude or deplore institutional treatment of individuals, or to add armchair support from the large and largely faceless public (Skoric 2012; Stieglitz and Dang-Xuan 2012). This free publicity has primed institutions to prioritize human factors in their policy and management, which has brought an unprecedented level of transparency into the daily workings of institutional social, ethical, and environmental agendas and constituents' daily lives.

In the efforts of policy makers and stakeholders to guarantee sustainable growth, stability, security, and progress, the struggle to find a common measurement variable is a common issue. Given its multi-dimensional structure, networked properties, and universality, well-being is well situated to be this variable (Kramer, Guillory, and Hancock 2014; J. Fowler and Christakis 2008; Hsee, Hastie, and Chen 2008; Huppert and So 2013). It is an underutilized yet effective concept for measuring populations' perceptions and expectations of themselves, services available to them, and their effects (Anderson et al. 2013). Well-being has been well-researched, and has shown reliable and robust measurements across time (Diener 1984a; Waterman 1993) making it more feasible to pursue than other normative, or values-based,

[2] http://www.worldbank.org/depweb/beyond/beyondco/beg_all.pdf. Last accessed: 10 March 2015.

assessments (Diener and Seligman 2004; Diener 2006). It is now being researched as a conceptual and practical complement to a myriad of macro and micro economic indicators, for mental health assessments, and as policy and decision making tools. Well-being has further attributes that make it attractive for institutional measurement. It is an overarching goal of both individuals and groups (Ryan and Deci 2001), making it intrinsically attractive to decision makers (Hsee, Hastie, and Chen 2008). Trivially stated: Everyone wants to be happier. Multiplier effects of high well-being include longer, healthier lives, and happier people are more productive and have lower absenteeism, leading to lower healthcare costs and turnover, and thus more favorable institutional reputations (Diener and Chan 2011; Vaillant 2008; Harter, Schmidt, and Keyes 2003). Well-being has been found to increase loyalty (Harter, Schmidt, and Keyes 2003) and has contagious network effects (J. Fowler and Christakis 2008; Kramer, Guillory, and Hancock 2014). Finally, experiencing well-being allows itself to be easily reported across digital mediums (Balahur and Hermida 2012). Given the centrality of digital presence in day to day life, specifically this factor reinforces the will of institutions to pursue well-being measurements in their interactions (Hall et al. 2012).

Due to the reasons alluded to above, societal well-being has become an overarching policy and management goal (Kahneman et al. 2004a; Stiglitz, Sen, and Fitoussi 2009). Creating decision scenarios where well-being is the goal and not the fringe benefit is complementary to a servitized, networked economy (Vargo 2009, 378). Institutions of every size, from state governments (Thinley 2011; Stiglitz, Sen, and Fitoussi 2009), companies (Harter, Schmidt, and Keyes 2003) to (digital) communities (White and Pettit 2004) are beginning to introduce well-being measurements in their decision making scenarios. However, this is still a relatively new phenomenon. Before 2000, well-being was not used as a management decision variable or policy instrument. One reason for this is measurability. Until recently, economic indices or macro social indicators (e.g., literacy rates, maternal survival rates) stood proxy for societal well-being. Due in part to the availability of ever more personalized, individual data sources (i.e., social media), these indicators are seen as no longer sufficient. Criticisms coalesce about the available indicators: they are one-dimensional as they are domain-specific, and refer only to very specific parts of progress without networking information into the context of wider developments (Stiglitz, Sen, and Fitoussi 2009; Veenhoven 1984; Auer 2011; Frey and Stutzer 2012). Especially the lack of networked information is a serious criticism. Furthermore, due to their methodology, such indicators highlight condition changes considerably after their occurrence. Again, in a digitalized economy, this is no longer sufficient. Finally, such measurements are also constrained by traditional aspects of scalability.

Well-being has been established as a valid and valuable indicator for progressive community management. However, despite its many attributes, institutions have been hesitant to implement a full-blown well-being measurement tool (White and Pettit 2004; Ahn et al. 2011). Known is that current indicators are restricted; consequently, institutions have been unable to use them as a comprehensive, detailed, and prompt institutional management service for stakeholders and policy makers. This leaves the open research challenge of designing a well-being indicator as a decision support service. The application of well-being as an indicator is occasioned by other questions viz., how can institutions discover how best to serve and engage their stakeholders? This, along with several concerns detailed below has been the stumbling block of progressive institutions in their efforts to implement well-being indicators in their decision making scenarios.

1.2 Research Challenges and Outline

Summarized, well-being must undergo a ***defining*** process by which it along with its data sources is satisfactorily and singularly demarcated; lest there be significant measurement issues. Once appropriate data sources have been identified, issues of data veracity come into play (***refining***). Finally, in order to use well-being as an institutional management service, stakeholders and policy makers must map perceptual states onto actionable items (***applying***), which is no trivial task.

Defining Well-being

Defining well-being is the foundational and essential first step in implementing well-being indicators. Adding constituent well-being to the assessment of broad social indicators requires that well-being (individual or communal) be defined in a way that is consistent and easy to measure, and in the best case with a framework in place to ease the making of normative judgments (Ahn et al. 2011; White and Pettit 2004). Since the 1970's psychologists and social scientists have worked at operationalizing well-being and its measurement instruments. By and large they have concentrated on two central themes: being happy, and being fulfilled (Ryan and Deci 2001), where happiness can be measured ordinally or cardinally (Frey and Stutzer 2001). While related, these aspects are not the same, with fundamentally different assumptions and indices of consideration (Dodge et al. 2012). The fundamental challenge until now has been the unsolved problem of isolating if well-being is experienced when one is feeling well, doing well, or attempting to be better (or, a combination thereof). As such, well-being lacks a *fil-rouge* and therefore a measurement instrument which leaves stakeholders unable to confidently apply well-being measurements for institutional management. If institutions aim to measure (or increase) constituent's well-being, this must be addressed. This thesis attempts to fill this void by addressing Research Question 1.1.

Research Question 1.1 ≺ Defining Well-being ≻ *Which attributes of well-being's conceptual definitions allow for the operational usage of well-being in institutional management?*

Well-being’s various definitions each have particular strengths allowing their application in an institutional setting. This fundamental split between the various definitions proposed in psychology has not yet resolved itself. This leaves institutional managers and policy makers underequipped with the necessary tools to define their measurements, thus unable and unwilling to further pursue well-being to support institutional management. In finding an operational definition of well-being, this thesis will contribute to the application of well-being as an institutional management service.

Increasing the well-being of individuals, and leadership capability to foster well-being organizationally co-creates the conditions necessary for healthy, happy institutions. Initial work on the integration of well-being and service design was proposed by (Rosenbaum et al. 2011; Anderson et al. 2013). Therein they propose (but do note validate) a framework to integrate consumer well-being and the service value chain. These contributions are broadly called Transformative Service Research (TSR). While a step in the right direction, the missing validation thereof means that the approach is lacking on several significant aspects required for functionality of such a framework.

Firstly, their framework is an entity map. As well-being is a normative state (White and Pettit 2004), interaction effects of the environmental and personal aspects on the service’s perception must be taken into consideration. Currently missing in the approach of existing literature, this is an important aspect. Also missing in this approach is granularity, meaning sub-community assessments and individuals’ perceptions’ of well-being are not in scope.

When considering implementing well-being as an indicator, the overarching goal in research and practice is gaining an understanding of the more nuanced and granular aspects of what it means to be a part of a community, and how individuals interact and feel about their community. Realized as a comprehensive well-being metric, it should be possible to build customizable reports based on community, sub-community, and/or constituent attributes which actively complement the attainment of personal, thus institutional, well-being. Design attributes include dynamic capabilities for institutions to monitor and track well-being, encourage stakeholder participation, and respond with appropriate policies. Such support mechanisms serve as a platform for testing alternative measures of well-being, and tracking changes in behavior and sentiment. Such requirements lend themselves well to being addressed in a service design framework. Thereby, the platform itself becomes a service for refining how well-being is measured. Considering the extension of TSR for progressive community management, this thesis next aims to answer the question:

Research Question 1.2 ⧼ Transformative Service Research ⧽ *What are the necessary attributes for constructing well-being oriented service design for institutional management?*

Continued and expanded research aimed at designing and realizing this as a structured computational tool (well-being oriented service design), processed in full depth and scope is necessary as it currently does not exist. Naturally, before a service is designed, its requirements must be identified and mapped: along with this are serious legal, organizational, and ethical implications deserving consideration before a well-being indicator is deployed. This makes mapping well-being to tangible policy and decision mechanisms non-trivial, requiring subjective assessment and policy management, as well as computational support. There is a need for standardized applications and user interfaces to deliver a higher quality of service, which assists decision makers in maintaining or increasing constituent well-being. Also necessary to address is once well-being data has been mapped to transformative services, what are the expected outcomes? This requires an assessment of how constituents are interacting to form a baseline. It also requires measurements on what if any differences occur. Further to the weaknesses of the current literature around TSR is the treatment and use of well-being data, which is not covered in previous works.

Well-being and its assessment are inevitably based on normative factors like values and judgment (White and Pettit 2004). In even the most homogenous communities, differences in experience, values, and desires can exist. Without considering the compacted interactions of services and constituents' environments and day-to-day activities, well-being and services cannot be fundamentally linked. Finally, intriguing work from (DeNeve and Cooper 1998) suggests that well-being has prediction potential. Assuming this is correct, well-being data should be able to estimate ex-ante the effects of institutional policy changes (Davies 1962), thereby supporting progressive community management. In response to these open challenges, Research Question 1.2 identifies the attributes necessary for the creation of Transformative Services in institutional management.

Refining Well-being Data Collection

Digitalization has led to several promising areas for data collection as proposed in the works (Vella, Johnson, and Hides 2013; Tov et al. 2013; Burke, Marlow, and Lento 2010). Many social media platforms provide interfaces that permit access to data produced by individuals, groups, and companies, or elicitation of further data. By accessing and analyzing this data, it is possible to construct rich information models to facilitate complex interdisciplinary research methodologies. It must be noted that individual responses as gained from surveys and interviews are social science ground truth. Traditionally the major method for well-being studies has been longitudinal surveys. Surveys do not allow for highly granular, frequent overviews of personal well-being. Another method that has been applied is interviews and focus groups (e.g., (Commission 2011; Bhutan 2012)). Interviews allow for highly granular, personal assessments of well-being, but are costly in terms of time and funding, and do not scale well.

In order to mitigate the well-known issues of incentivization of participants and high costs researchers have proposed two mechanisms; serious games, and unobtrusive measurement (Deterding et al. 2011; Vella, Johnson, and Hides 2013; Deterding 2011; Balahur and Hermida 2012; Tov et al. 2013). When trying to circumvent the costs and possible bias accumulated in these methods, several rounds of calibration and verification are required. Here, computational support becomes necessary. Furthering data collection by adding HCI elements affords creation of the well-being maps of communities and/or institutions necessary to evaluate TSR (Mitchell et al. 2013). Such maps can be used to establish then track the general mood of a given population; they can also serve as an ex-ante measurement of changes from policy implementation (Dodds et al. 2011). HCI interfaces for mapping and design of an institutional well-being data collection and evaluation tool is a natural next step for policy making bodies and stakeholders in community management.

The open design and research challenge is harvesting well-being data:

1) Frequently,

2) At a low researcher-participant cost,

3) Which does not lead to participation fatigue.

Considering frequency, an issue to consider is that if asked the same question multiple times, participants may become disengaged or drop out of the study. Especially worthy of further investigation with respect to this are participation and truthful reporting. Participants may become disincentivized to continue participating with repetitious questioning; they may also report untruthful data for reasons ranging from disengagement to gamified personas.

Facebook is a particularly interesting platform for launching a TSR application due to its market share and structure. Facebook is the world largest social network and social media platform, consisting of 1.44 billion monthly active users.[3] This means that data is abundant and readily available. As opposed to other networks (e.g., Twitter, google+), Facebook allows full data feeds, assuming authentication rights are in place. However, Facebook's Application Programming Interface (API) and its Terms and Conditions have historically been less accessible to scholarly research unless conducted in-house. Accessing individual data streams outside of Facebook's research team required an app which crawled the data from the participant's profile (e.g., (Youyou, Kosinski, and Stillwell 2015; Schwartz et al. 2013; Catanese et al. 2011)) or requires frequent data input (Killingsworth and Gilbert 2010). This caused the situation of most Facebook research outside of its proprietary research office being completed qualitatively (Wilson, Gosling, and Graham 2012). Advances have since been

[3] http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/. Last accessed: 5 May 2015.

made to Facebook's Graph API and Terms and Conditions, lowering barriers to the data held within. While a well-positioned platform for the introduction of a TSR application, further research into the extraction methods and the impacts of said methods must be completed. This leads to Research Question 2.1:

RESEARCH QUESTION 2.1 ⪡ DATA HARVESTING ⪢ *Considering the methods gamification and text analytics, which is more appropriate for extracting near to real time well-being data from online social media in a continuous manner?*

Research Question 2.1 deals with two quite specific data extraction methods. Numerous methods, too many to be listed here exhaustively, exist and could be implemented. However, gamification and text analytics have particular traits that lend themselves well to the design and implementation of a comprehensive TSR application. Comparing gamification and text analytics allows for a comparison of stated preferences (gamified surveys) to revealed preference (sentiment analysis) with respect to the expression of well-being. Both methods lend themselves to the environment of Facebook, and each represents a (relatively) uncomplicated mechanism that stakeholders and policy makers could implement, considering a successful outcome.

These novel solutions are promising but need to address several public criticisms and challenges to validity; also the parameters of the two methods must be established. The gamification of survey mechanisms is promising but untested. It is assumed (but not proven) to have a motivational effect on participants in a variety of institutional contexts (e.g., education, corporate, physical health). Also unknown is how the interaction between participants and the survey changes when gamified, as well as if there are any impacts on participation. These open questions are addressed subsequently in Research Question 2.2.

The feasibility of extracting text from various sources depends on several factors, including identification of a community, veracity, 'noise' levels and technical scalability. Text analytics and its related methodology sentiment analysis have several public criticisms about the deficiencies, non-robust precision and recall, dependencies on frequencies or curated dictionaries, and inability to identify alternative meanings from text (Jungherr, Jürgens, and Schön 2011; Chung and Mustafaraj 2010). Another major research gap being currently addressed is the alignment and validation of (traditional) psychometric measures to this relatively new data source. Still missing are replicable studies and algorithms that unobtrusively (in an unobserved manner) collect, analyze, and report on this type of unstructured data. These open research challenges are addressed by Research Question 2.3.

Today we are habituated to maintaining our digital profiles and reveal more information about ourselves than ever before, laying convenient foundations for analyzing specific aspects of digital communities. This orientation allows for unprecedented access to highly granular, personal data that was before this untouchable in frequent intervals. Research Question 2.1 is posed as a comparative assessment of pulling responses from participants (gamification) and the reception of pushed data from participants (text analytics).

An emergent proposal is furthering current applications of human-computer interaction (HCI) to well-being measurement. Gamification is one such mechanism. Considering a TSR application, gamification's positive attributes include motivation, engagement, and excitement. Participants must be incentivized to answer questions frequently and truthfully. Participant motivation and engagement are critical success indicators for gamified well-being measures: without an active, engaged community pushing data into the system, this method cannot be applied in a large scale application. Excitement is necessary not only for making otherwise 'boring' tasks like survey completion interesting, but also to further network propagation. As such a system is envisioned to be 'opt-in,' network propagation is also critical for the success of the application. Finally, truthful, non-gamified responses are also critical to the output of such a TSR application. If this application is driven by anything other than honest well-being reporting, the system is not meritorious to be scaled up as a general community tool.

In a novel application of two before-unconnected aspects, certain foundational questions on suitability must be first addressed. It cannot be stated what serious games yields both continued participation and truthful self-reporting without first assessing if adding gamification to well-being data collection has a motivational effect on continued use. Corollary to that, a metric of truthful reporting must be benchmarked against existing literature to establish if participants are incentivized to answer truthfully when adding gamification mechanisms. As this is a layered problem, an iterative design solution is best applied to address Research Question 2.2.

RESEARCH QUESTION 2.2 ≺ GAMIFIED SURVEYS ≻ *Does the gamification of surveys enable frequent, granular views of individual's well-being without a high participant drop-out rate?*

Context-dependency of gamification methods is a best practice in the literature surrounding gamification and serious games. Implicit in this best practice is that new solution concepts and proof of concept applications must be iteratively modeled and constructed in order to adequately test the method's instantiation. This suggests that the gamification of well-being requires a tiered approach in order to properly evaluate the merits of the approach. Accordingly, RQ 2.2 is addressed in an iterative fashion.

The implications of gamified well-being data extraction are further into the domain of gamification and its applicability to well-being measuring. Being a current trend, there is a lively discussion on gamification that not only includes its definition and scope but, to some extent, also questions its fundamental suitability. Another contribution is the creation and evaluation of an innovative informative-driven solution. The release, spreading and technical evaluation processes are a relevant building block for evaluation of future, similar technical solutions. The findings revealed are poised to provide a valuable contribution to the further development of gamified well-being measuring.

Text-based data provides data that replicates revealed preferences research designs (and thus actual behavior), can be collected at any time, is abundant (in the era of social media), and is relatively inexpensive, a direct contrast to surveys and interviews. As such, it is being investigated as a related or replacement method for such time and cost intensive research designs. Methods like surveys and interviews are long established, and their strengths, weaknesses and common pitfalls are well-known. In the terms of surveys and interviews, the pitfalls are generally fall under the domain Common Method Bias (CMB). CMB and its remedies have been well-published and are well-regarded (Podsakoff et al. 2003; Conway and Lance 2010). This same process is currently a lacunae of digital research, where authors are only beginning to address bias and common pitfalls of data gathered on the internet and across different platforms (Zimmer 2010; Ruths and Pfeffer 2014; González-Bailón et al. 2014).

As cautioned and proven in a growing body of work (e.g., J. Chung and Mustafaraj 2010; Jungherr, Jürgens, and Schön 2011), analyses and results based on data which hasn't been properly treated must be taken with a grain of salt. However, the parameters of data preparation for unstructured data are still emerging. This leaves considerable room for both the development of standards, and for poorly designed research to receive unnecessary attention (cf. Jungherr, Jürgens, and Schön 2011; Tumasjan et al. 2010; cf. Wang et al. 2014; Kramer 2010). Looking more carefully at the application of unstructured textual data to the assessment of individual well-being, open questions remain on the alignment of individuals' survey responses and their self-produced text as extracted from the platform Facebook.

The results of psychometrics surveys are considered to be representative of actual personality. To be established are the suitability of text in making psychometric assessments, along with an appropriate method to validly and reliably extract these traits. Also, which features are available from text and latent sentiment to robustly represent these traits? These questions are pertinent both from the perspective of moving the TSR agenda, as well as from the validation of different analytics methods on different online social platforms. Research Question 2.3 establishes the relationships between self-produced text and survey responses.

RESEARCH QUESTION 2.3 ⊰ RELIABILITY AND VALIDITY ⊱ *Which well-known relationships between well-being and personality can be reproduced when using text-based data found in social media posts?*

First research has been done (Kramer 2010; Burke, Marlow, and Lento 2010), but the method has been heavily criticized in the works (Wang et al. 2014; Jungherr, Jürgens, and Schön 2011) for the concentration on single-item indicators (mentions of words like happy, sad), the lack of context sensitivity, and the weight given to term frequency. The output of text analytics tools is per definition arranged according to the higher logic of the program or algorithm applied in order to (re)structure the newly structured data. Thereby words and phrases can be sorted, placed, and assessed. Such categories have unknown latent relationships to the items of surveyed psychometric tests. Fully parameterizing these latent relationships for a given method-platform pairing is necessary for the utilization of unstructured text and its analysis methods (text analytics and sentiment analysis). Once these assessments are made and properly evaluated for the Facebook scenario, community analysis as well as individual personality and well-being can be fed into a full-blown TSR application.

Further challenges face scholars applying online gathered social data generalizable social models. Digital anonymity can enable gamified personas, presentation of idealized self(s), or even online disinhibitive behaviors (trolling) in the most extreme cases (Hilsen and Helvik 2012; Ellison, Heino, and Gibbs 2006; Buckels, Trapnell, and Paulhus 2014). These are also the overt cases of actively altered personalities (Lingel, Naaman, and boyd 2014), which is nearly untouched in research. Specific to gamifed scenarios, a danger of gamified selves can occur given the playful environment being introduced (Dixon 2011). Even more than active (mis)representation, it is necessary to consider is if the same person alters their personality based on the constraints of the platform in use (Davenport et al. 2014; Lin and Qiu 2013). When individuals can create idealized selves without a cross-validation of actual personality, data veracity is of the upmost concern (Back et al. 2010; Caspi and Gorsky 2006; Utz 2005; J. Hancock 2007).

Pertinent questions on idealized self and its treatment in data handling are: the verification of data gained on social networks to actual personality, and appropriate uses in community management and policy-making. Considering the two scenarios introduced in Research Question 2.1, this takes two very different forms. In the scenario of serious games, the core consideration here is the designing of an incentive or motivation scheme that encourages participants to push truthful responses about their state of mind into the system. In text-based scenarios, first the relationship between self-reports on surveys and self-produced text considering the use case of Facebook must be established in order to find out what extent is it possible to use self-produced text to diagnose deceptive profiles. This leads to Research Question 2.4.

RESEARCH QUESTION 2.4 ⊰ DATA VERACITY ⊱ *Are discernable characteristics of active representation identifiable, and if so, what are these characteristics?*

What has not been approached in a systematic way is the verification of such data on offline and actual personality. Worrisome is the near inability of the researcher to verify that data extracted from online social networks and online social media aligns with actual people and their real life thoughts, concerns, and personalities. From this perspective, analyses based on online social media re promising due to their broad reach and appear, but risk lacking veracity necessary to build generalizable social models. This is a research gap that must be addressed.

Scholars in the social sciences and computer science have not yet adequately addressed controlling for what can be called self-representation, or the propensity to display or censor oneself, in their analyses (Zhao, Grasmuck, and Martin 2008; Das and Kramer 2013). Research Question 2.4 is at once a design aspect as well as a data management aspect. Positive results in accordance with this question support the creation of a best practice standard of mitigating bias in online social media data.

Applying Well-being Measurements

Granular, localized information can be unobtrusively gathered to assess indicators of well-being. This information is already abundant and available via online social media. The missing link is a rigorous, anonymized and open source artefact that gives feedback to stakeholders and constituents. Necessary for these research goals are the mapping of communal characteristics. This thesis addresses this research gap by addressing each of the listed research questions subsequently. The final step is the realization of a full blown TSR application, considering the findings of each phase of the research. The realization thereof is an empirical demonstration of well-being's applicability and validity as a progressive community indicator.

Summarized, necessary questions to be addressed in a successful demonstration include:

- Considering the operationalized definition of well-being established in Research Question 1.1, what is required to identify communal well-being from online social media data?
- Which features identify an emotive baseline of communal discourse?
- Do changes in sentiment identify major events within a community network? If so, what are the requirements for such tracking mechanisms?

These characteristics form the baseline from which to identify and measure the quantified attributes of communal well-being. Accordingly, these aspects must be address in future community modeling and prediction works. Research Question 3 in its full depth and breadth addresses the identification of communal characteristics via sentiment analysis and context-sensitive text mining. This research question addresses the noted criticisms of text analytics by applying broad sentiment analyses as opposed to positive and negative emotion analyses. In support of this effort, the following Research Question is addressed:

RESEARCH QUESTION 3 ⊰ CHARACTERISTIC MAPPSING ⊱ *Can community characteristics like well-being and organizational belongingness be unobtrusively established? If so, what are the key characteristics?*

RQ 3’s intended contribution is event-based tracking from online social media data. This is interesting from a policy perspective, as it creates a communication mechanism for where stakeholders can present and discuss events and policy changes in a public forum. It is also a positive demonstration of the usefulness of progressive community management by the way of Transformative Service Research.

Having first established the requirements and design aspects necessary for such a tool, this thesis’s contribution is a valid TSR application from which to make community modeling and predictive assessments. Developing technology-enabled services to improve well-being is named as a strategic priority of service science in the 2015 Journal of Service Research ‘Service Research Priorities’ article (Ostrom et al. 2015, 140). A successful completion to this thesis fulfills the research gap of a valid, empirical, information-driven TSR application.

1.3 Thesis Structure

The research outline presented in the previous section reflects the structure of this thesis, which encompasses four parts. Part I introduces the research questions and development, as well as use cases. Part II discusses foundations, seminal terminology, and lays out the applied methodologies, which are addressed in Part III. The evaluation of the methods in their varied use cases are also encased in Part III. Part IV concludes the thesis and highlights future research directions. A high-level illustration of this work’s structure is shown in Figure 1.1.

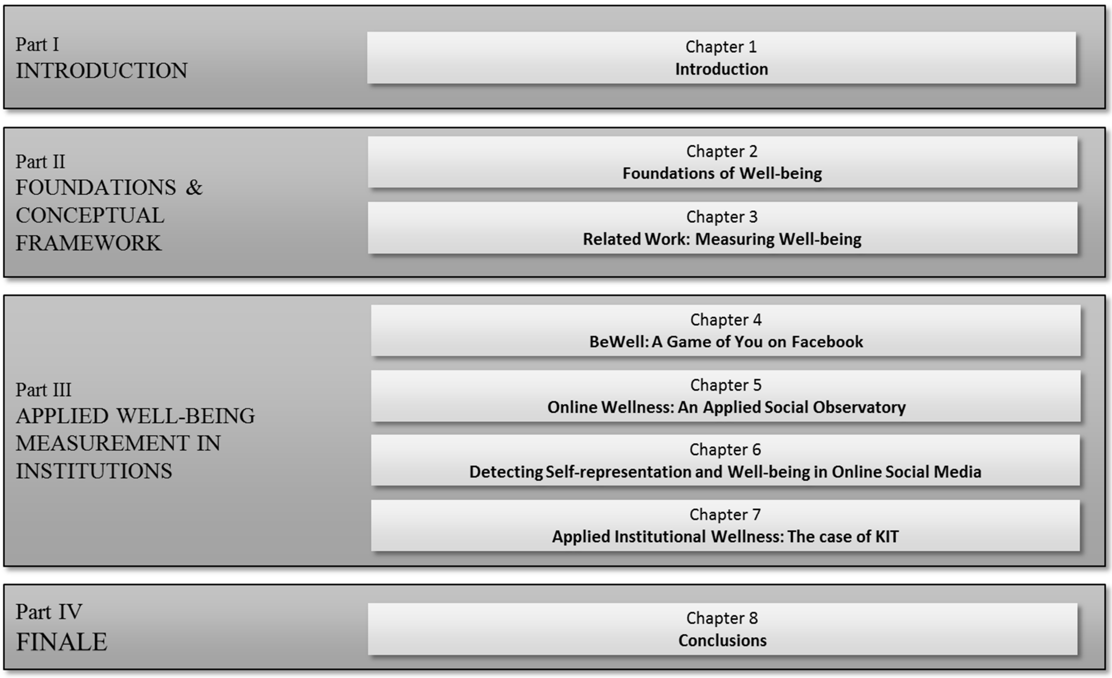


Figure 1.1 Thesis structure

Chapter 2 introduces the formal descriptions of well-being, discusses existing literature and the state of the art measurements of well-being, and proposes a working definition of well-being measurement for the purposes of this research. Chapter 3 lays down the foundations for the research's approach by introducing a structured framework for the analysis of well-being measurement. Existing efforts in the quantification of well-being and data sources are addressed. Additionally, two promising methodologies for the measurement and detection of well-being, namely gamification and text analytics, are presented.

Chapter 4 is the first of four case studies applying the framework and methods from Chapter 3. Specifically, this chapter discusses the application of gamification to the surveys discussed in Chapter 2 to incentivize use participation. The written expression of emotion is the basis of the rest of the thesis. Chapter 5 introduces and validates the use of text analytics as a mechanism to detect sentiment in and of online communities. Chapter 6 discusses the implications of online personas in the use of online social media data in research design, and suggests mechanisms to minimize this type of participant-introduced bias. Building on this, Chapter 7 combines the implications of Chapters 5 and 6, and assess the well-being of a university campus based on their Facebook presence. Chapter 8 summarizes the key contributions of this thesis, provides an outlook on future research, and highlights complementary research topics.

1.4 Research Development

Parts of this work have been developed and published in peer-reviewed international conferences and international journals. This section discusses the outlets, development of work, and subsequent extensions contained in these chapters. Moreover, the main contributions of the research and their integration into current research projects are highlighted.

Part II considers defining then operationalizing well-being in a service ecosystem. Initial discussions on the integration of well-being and service design were presented at the American Marketing Association's Special Interest Group on Services (ServSIG) conference (Hall et al. 2014). This paper discusses the formalization of a well-being measurement system in accordance with TSR principles, highlights data lacunae, and introduces the argument that considers when human happiness is at stake, more doesn't always signify better. Foundational to this paper is the need to move from a theoretical standpoint to applied transformative service research. Not only does this work set the stage for the theoretical contribution in Chapters 2 and 3 (Foundations and Related Work), it has also begot two applied service research studies: service zone design as a tool for public good in the case of food deserts (Johann et al. 2014) and service requirements for citizen participation in the German national legislative action *Energiewende* (Energy Transformation) (Bertsch et al. 2015).

Part III discusses two applied research methods for operationalizing well-being: gamified surveys, and text analysis. The development and evaluation of these two methods have been published in the proceedings of one workshop and three conferences, *as well as two international journals*. The initial proposal to gamify the survey items of well-being measurement was published at the 2012 Analyzing and Improving Collaborative eScience with Social Networks workshop (eSoN 12) (Hall et al. 2012). The implications of this proof of concept work are twofold: an incentivization scheme is necessary for continued participation, and that alternative methods of well-being measurement (text analytics) may be put to use in order to use well-being as a predictive indicator. Gamified incentives and Facebook-oriented participation patterns are reviewed and extended in the work (Hall et al. 2013), which was presented at the 2013 Social Computing and its Applications conference (SCA13). A major finding of this work is the role of personality in individual well-being assessment. The work (Hall, Caton, and Weinhardt 2013) confirms the previous works' personality finding, introduces longitudinal assessments of personal well-being, and discusses the potential for machine learning to replace standard analysis packages in well-being evaluation. This work was presented at the Human-Computer Interaction International (HCII) conference in 2013. An extension of (Hall, Caton, and Weinhardt 2013) compares the performance attributes of machine learning algorithms when predicting well-being scores based on real data (Wilckens and Hall 2015).

A common (and rather well known) limitation of surveys discussed in the above works is respondent bias. Specifically, reference effects and selection bias cannot be estimated in online social media environments. A novel mechanism discussed in (Hall et al. 2012) is the application of text analytics tools for the estimation of well-being. Another operationalization of Part III introduces exactly this, in the form of the journal article (Caton, Hall, and Weinhardt 2015) *in Big Data & Society*. This article presents unstructured text from communal discourse as a progressive indicator of happy societies, with the use case of German politicians and their Facebook followers. The implications of this article are that sentiment analysis is a valid and replicable method to estimate community discourse, and that the original language (German) must not be altered to English for good performance. This article has been extended for the thesis by an in-depth description of the extractor's architecture and functionality.

A research challenge identified in (Caton, Hall, and Weinhardt 2015) is the lack of ground truth in unobtrusively gathered social media studies. Chapter 6 of Part III addresses this challenge. (Hall and Caton 2014), a preliminary review of the results, was presented at the Oxford Internet Institute's symposium on Internet, Policy & Politics (IPP2014). Insights of this work are the basis of the chapter, which finds participants misrepresent their own writings, leading to participant bias in cases of unobtrusive research designs. The full evaluation of this study has not been published elsewhere.

The final chapter of Part III is a compilation of the findings of Chapters 5 and 6. Chapter 7 focuses on the Facebook community surrounding the Karlsruhe Institute of Technology. Therein, it first isolates the self-representation bias as proposed in Chapter 6, and then applies communal discourse methods from Chapters 5 and 6 to assess the well-being of the KIT community. Considering the evaluation of these methods, a research-in-progress work was accepted by the ACM Factors in Human Computing (CHI2015) conference (Lindner et al. 2015), where a subset of the data was presented and discussed as a proof of concept work.

Part II.
Foundations
and
Conceptual Framework

Chapter II Foundations of Well-being

"Human well-being is not a random phenomenon. It depends on many factors - ranging from genetics and neurobiology to sociology and economics. But, clearly, there are scientific truths to be known about how we can flourish in this world."

"The Moral Landscape: How Science Can Determine Human Values," (Harris 2010)

People and institutions that are flourishing share certain characteristics: higher productivity, learning that is more effective, more stable social ties, and better health and life expectancies (Huppert and So 2009; Grawitch, Gottschalk, and Munz 2006; Smith Warner 2013; Frey and Gallus 2013; Diener and Chan 2011). High well-being inter alia supports "effective learning, productivity and creativity, good relationships, pro-social behavior, and good health and life expectancy" (Huppert and So 2013). This creates multiplier benefits for society: higher well-being can contribute to less expenditure on programming curbing social disintegration, lower healthcare costs, lower absenteeism, and overall "performance" increases (NEF 2009; Gasper 2005; Oishi, Diener, and Lucas 2007; Harter, Schmidt, and Keyes 2003). This chapter addresses key conversations in the scholarly literature in well-being measurement, framing the interdisciplinary understandings of well-being for use in institutional management.

2.1 Towards an Interdisciplinary Definition of Well-being

Well-being is evaluated in a variety of ways: as subjective well-being, psychological well-being, or via economic calculation (Diener et al. 1999; Diener 1984a; Waterman 1993; Waterman, Schwartz, and Conti 2006; Samman 2007; Ryan and Deci 2001; Karlsson, Loewenstein, and McCafferty 2004; Zamagni 2014; Stiglitz, Sen, and Fitoussi 2009). While each domain has different strengths, when used as complimentary systems they create a fitting proxy of individual and institutional well-being (Samman 2007; Huppert and So 2013; Gasper 2005). There are two major literature strains based in philosophy and psychology covering the concepts of well-being: one on hedonic well-being (Diener 1984b; Diener 1984a; Diener and Suh 1997), the other on eudemonic well-being (Ryan and Deci 2001; Huppert and So 2013; Ryff and Singer 2013). The distinction is also labeled subjective well-being (SWB) versus psychological well-being (PWB) in the literature. This work uses the terminology interchangeably. The psychological field of study is known as "positive psychology." The coming section defines SWB and its measurements, and is followed with a discussion of eudemonia's varying definitions and measurements.

2.1.1 Economic Assessments of Well-being

Economic assessments of well-being equate tangible measurements like income, wealth, social security and safety with well-being. It is based on the assumption that certain levels of these economic measures allow individuals to achieve personal fulfillment, which again results in well-being. Economic perspectives of well-being are popular, since it is relatively easy to measure, tangible, and widely used in support of political decision making (Frey and Stutzer 2012; Frey and Stutzer 2001; Diener and Suh 1997; Ahn et al. 2011). However, in the transition to indices of revealed preferences (ordinal utility) as the gold standard of behavioral and choice measurement in microeconomics (Robbins 1932), cardinal utility, such as that found in cost-benefit analyses, has fallen into disuse. Cardinal preference is however paramount to the measurement of well-being as it is commonly collected today. Accordingly, as interest in economic psychology increased in the past decades, works applying cardinal measurements of well-being and happiness have increased (Frey and Stutzer 2007; Frey and Stutzer 2012; Kahneman 2009; Kahneman and Thaler 2006). Well-being in the economic sense has been formalized by (Frey and Stutzer 2001, 30–31) as the following function:

$$W = H[U(Y,t)] + \epsilon \tag{2.1}$$

where *W* represents self-reported well-being levels, generally obtained via a Likert scale (i.e., the Satisfaction with Life Scale (Kahneman et al. 2004b)), and is thus cardinally bound. The function *U(..)* denotes well-being (in the sense that well-being is measured as a utility function), and Y is the determinate set of the respondent's reported well-being. *t* indicates that the relationship between Y and U can vary. The continuous non-differential function H[.] relates well-being reports and actual well-being, where H[.] rises if U increases. The error term ϵ relates to the relationship between actual and reported well-being by capturing latent variables that impact well-being reporting.

Economic well-being measures are not intended to provide insights about personal well-being levels, but about well-being on a more general, averaged, or national basis. Foundational economic theorists including Adam Smith and Jeremy Bentham recognized the limits of using income and material wealth as the sole definition of economic utility (Smith 1776; Bentham 1789). Nevertheless, several studies support a correlation between economic well-being and SWB on a macroeconomic scale (Stevenson and Wolfers 2008). (Diener and Seligman 2004) explain the importance of economic measures for well-being particularly for the "early stages of economic development, when the fulfillment of basic needs was the main issue" (p. 1), but relativize this importance for highly developed countries. This assessment is based on what has been defined as the 'Easterlin Paradox,' which describes a saturation point in the relationship between income and well-being on a national basis (Easterlin 1974). Easterlin's original argument was that happiness increases with income in developing countries. However, after a saturation point of income is hit ($10,000), well-being and income no longer have a positive significant relationship, but rather a negative relationship. The finding has been confirmed several times (Easterlin 1995; Blanchflower and Oswald 2008; Easterlin 1974; Kahneman et al. 2006) and is not only observed in comparisons between countries, but also in time-series analyses for averaged national data. Economically saturated countries, e.g. the United States, do not obtain higher averaged well-being when the income per capita rises over time (Clark, Frijters, and Shields 2008). The paradox is explained by decreasing importance of additional income once basic needs have been satisfied (Stevenson and Wolfers 2008). This argumentation is however debated, with other economists reporting different findings (Stevenson and Wolfers 2008; Gasper 2005; Preziosi 2013). These studies however tend to be smaller, and are less widely accepted for methodological reasons (Easterlin et al. 2010). Nevertheless, until now policy decision making is mainly still based on the underlying idea of economic well-being that increased wealth and social status lead to higher well-being within the society. Economic well-being is therefore widely used as an argument in favor of economically beneficial development (Gasper 2005; Kahneman and Krueger 2006).

2.1.2 Philosophical and Psychological Foundations of Well-being

What does it take to be well? There is a general overlap between the two notions of well-being, though interestingly, these two definitions can also have conflicting outcomes. Both tend to consider overall satisfaction with life as a necessary metric for the existence of a good life, considering both an individual person and/or a community (Veenhoven 1984; Veenhoven 2010; Veenhoven 2013). Where SBW estimates temporal feelings of happiness, PWB concentrates on the process of setting, striving for, and attaining self-betterment goals. This is a critical difference, as the measurement system in place dictates the outcomes when considering well-being as an indicator for progressive community management.

The major philosophical foundation of hedonistic well-being is that the goal of life should be to experience the maximum amount of pleasure, as the pursuit of happiness is the ultimate goal of life. Happiness is found when one is pleased; it does not mean that whatever pleases a person is enriching or good for them. One can be happy without being (mentally, emotionally, or physically) well. SWB is the "happiness" (or hedonistic) side of the well/being argument (Diener, 1991). This is best crystalized in the argumentation on the good life by philosophers like Aristippus, Hobbes, and DeSade, who saw the major goal of life through the lens of satiation of human appetite, pleasure, and happiness (Ryan and Deci 2001).

Eudemonia is the attainment of the self, occurring when life activities are meshing with one's most deeply held values (Waterman, 1992). The things which make one happy and the conditions which makes one thrive are not necessarily the same; temporal instances of feeling good (happiness) are not necessary to achieve well-being. This is the inverse of SWB: one can achieve well-being without being happy about it. This is a view advocated by foundational philosophers like Aristotle and Fromm. Artistole in fact considered the pursuit of happiness to be vulgar, as individuals should be elevated above the slavish pursuit of desire (Ryan and Deci 2001). The debate between happiness and eudemonia and its place in the attainment of well-being has lasted millennia and centers around the ideas of happiness versus satisfaction, introduced in the coming sections.

Happiness is a Warm Gun: Subjective Well-Being

Subjective well-being, the most widely researched aspect of well-being, is an indispensable component of positive psychological health, although is not a sufficient condition for it (Ryan and Deci 2001; Frey and Stutzer 2001). While the first attempts to define SWB rather looked into demographics (W. Wilson 1967) or socio-economic status (Easterlin 1974; Easterlin 1995), other researchers (most notably the works of Diener and colleagues) tried to have a closer look into the components of SWB and their interactions and tried to give a greater recognition of the central role played by people’s goals, coping efforts, and dispositions (Diener 1984b; Diener 1984a; Pavot and Diener 1993; Diener et al. 1999).

SWB surveys one’s total life satisfaction, the presence of well-being, and the absence of negative feelings (Lyubomirsky, King, and Diener 2005; Diener 1994; Diener and Suh 1997; Diener 1984a). Purposefully absent of objective conditions such as health, comfort, virtue, or wealth, SWB looks solely at one’s assessment of their state of life (Kahneman and Krueger 2006; Kahneman et al. 2004b). Although such factors are potential influences on SWB, they are not seen as an inherent and necessary part of it (Diener 1984a, 543). The exclusion of objective conditions allows for a comparison of the well-being levels of persons with quite different living conditions, facilitating wide applicability of SWB. However, it is reflective in nature, meaning that assessments of well-being are necessarily backwards-looking.

A characterizing feature of SWB is the inclusion of positive and negative affect (emotions), which means the pure absence of negative factors does not constitute high SWB. This distinguishes SWB from most measures of mental health where the focus is laid predominantly on negative measures of well-being (Huppert and So 2013; Diener 1994). The most commonly used scale to assess SWB is the Satisfaction with Life Scale (Diener et al. 1985; Pavot and Diener 1993; Frey and Gallus 2013), measured as a five item, seven-point Likert scale. The score is the mean of the five items. (Diener et al. 1985) claims that single item measures are temporally less reliable than multi-item scales. They can be more susceptible to types of so-called acquiescence response bias where participants tend to agree with all items, and most significantly, are subject to being invalidated by poor wording.

____ In most ways my life is close to my ideal.

____ The conditions of my life are excellent.

____ I am satisfied with my life.

____ So far I have gotten the important things I want in life.

____ If I could live my life over, I would change almost nothing.

Another technique addresses the problem of biased information with a close link of the question to a certain event or activity. The “Day Reconstruction Method” (DRM) by Nobel Prize winning researcher Daniel Kahneman and colleagues identifies the remembered well-being for each activity and experience of the preceding day (Kahneman et al. 2004b). The participants “first revive memories of the previous day by constructing a diary consisting of a sequence of episodes. Then they describe each episode by answering questions about the situation and about the feelings that they experienced” (Kahneman et al. 2004b, 1776). The review of the previous day causes that recent memories lose dominance, so that errors and biases of recall are reduced (Kahneman et al. 2004b). The survey part of the method is based on the experience sampling method (ESM) (Scollon, Kim-Prieto, and Diener 2003), as feelings in different situations are aggregated towards an overall well-being measure. But deviating from the ESM, (Kahneman et al. 2004b) propose that the DRM allows for measuring a sufficient number of different events during just one day as well as enough days in a time series and is therefore more efficient.

Although well-established, criticisms of dimensionality and possible biases of SWB are still plentiful (a good overview is found in (Angner 2005)). This encourages cross-disciplinary scholars to extend the definition and measurement of SWB with even more cutting-edge and validated methods. Especially (Frey and Stutzer 2012; Frey and Stutzer 2001) argue that not only subjective but also objective measurements of happiness are necessary. Figure 2.1 illustrates Frey and Stutzer’s proposed continuum of happiness measurements, including physical and neurological assessment, Kahneman’s sampling method, as well as the Satisfaction with Life Scale. It is important to note that as classical economists, Frey and Stutzer proposed but did not validate physiological and neurological measurements.

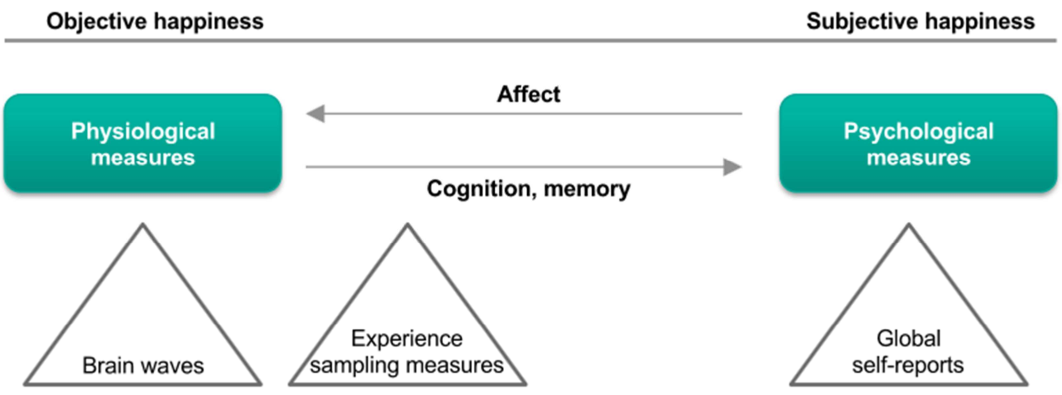


Figure 2.1 Frey and Stutzer's proposed continuum of happiness measurement

The validation of objective happiness and thereby SWB as expressed in Figure 2.1 is an on/going research area. (Rutledge et al. 2014) proposed the closest representation to date of a formal expression of (objective) momentary happiness in a gambling experiment with Neuro-Information Systems, establishing this function across (n=18,420) participants:

$$Happiness(t) = w_0 + w_1 + \sum_{j=1}^{t} \gamma^{t-j} CR_j + w_2 \sum_{j=1}^{t} \gamma^{t-j} EV_j + w_3 \sum_{j=1}^{t} \gamma^{t-j} RPE_j \quad (2.2)$$

where *CR* is a certain reward, *EV* is the expected value of an action, and *RPE* is the difference between expected and actual rewards. *t* is the moment of assessment, w_0 is a constant term, other weights *w* capture the influence of different event types. $0 \leq \gamma \leq 1$ is a forgetting factor that makes recent events more influential than those before. CR_j is the *CR* if chosen instead of a gamble at the time point *j*. EV_j is average reward of the gamble if chosen at the time point *j*, and RPE_j is the *RPE* on trial $_j$ contingent on choice of the gamble. If the *CR* was chosen, then $EV_j = 0$ and $RPE_j = 0$; if the gamble was chosen, then $CR_j = 0$. They established that momentary happiness is not a response to outcomes of a reward-based task based on current earnings, but rather the combined influence of recent reward expectations and prediction errors arising from said expectations (Rutledge et al. 2014, 1). In addition to showing the link between mental processes and happiness, this study provides an important clue into the nature of momentary effects on one's overall happiness.

Eudemonia: A Structured Diversity of Joys

Even with successive attempts to define well-being, quality of life, and happiness, there is still no consensus definition of eudemonia (Varelius 2013; Veenhoven 2013). Eudemonism is more diverse and considered by some a more sophisticated well-being measurement system (Waterman 1993; Ryff 1989; Page and Vella-Brodrick 2008; Ryff and Keyes 1995). In contrast to SWB, these scales are not only about general life satisfaction (Samman 2007). Rather, they consider factors that influence ones' inner self-fulfillment and inner growth (Waterman 2007; Waterman, Schwartz, and Conti 2006). The central goal is the actualization of one's self in order to thrive and grow (Waterman 1990). Generally self-actualization is pro-social, and can be pursued and experienced in the present and future tenses. Being pro-social and forward looking allows PWB to be considered in efforts to design a well-being based community management system and related policy mechanisms (Ahn et al. 2011; Veenhoven 2008). However, eudemonism fails to coalesce into a single, widely used scale due to its wide reaching scope and failure to agree on minimally required measurable items. Moreover (Veenhoven 1984) suggests to include "non-verbal cues" (p. 46) and "expert ratings" (p. 47) into the assessment. While expert ratings are questionable, as only the individuals verify how happy they are, non-verbal assessments like those found in self-produced text are addressed in the coming chapter (Section 3.2.3).

In order to make eudemonic measurement feasible, various PWB scales have been developed (Ryan and Deci 2001; Ryff 1989; Ryff and Keyes 1995; Hsee, Hastie, and Chen 2008). Generally, areas surveyed by PWB instruments consider domains like autonomy, environmental mastery, personal growth, positive relations, purpose in life, and self-acceptance (Ryff and Keyes 1995; Ryan and Deci 2001). Such criteria are considered to illustrate the extent to which one is accomplishing basic psychological needs. Fulfilling these will result in better health, both physical and mental, thus amplifying PWB. PWB too suffers from the criticism that it is highly subjective; that is to say, the individual sets and assesses their individual criteria (Samman 2007). Criticisms of SWB's subjectivity notwithstanding, it is important that all those factors are measured by people on their own scale; that the goals are set by themselves; are guided by their wants; and each domain is only fulfilled up to a degree that they feel comfortable with (Ryan and Deci 2001). Such a process leads to self-actualized individuals and communities, which are healthier and happier individuals and communities. In contrast to Diener's Satisfaction with Life Scales, PWB scales are frequently single-item, as single-item scales have been found to perform just as well as multi-item scales in the case of clearly worded items (Bergkvist and Rossiter 2007).

Self-Determination Theory

Self Determination Theory (SDT) is one of the most widely used extensions of eudemonic theory, as it lends itself nicely to public policy and institutional goals of increasing public well-being, without complete reliance on the subjective assessment of the individual. (Hirschauer, Lehberger, and Musshoff 2014; Ahn et al. 2011; Veenhoven 2008; Frey and Stutzer 2007). It sets personal well-being not only equal to self-fulfillment, but also considers the basis that has to exist in order to achieve well/being or pro-social goals. This basis consists of self-determination, competence, and relationships with others (Ryan and Deci 2001; Vella and Johnson 2012). Self-determination is the feeling of empowerment to follow one's own decisions and act on their own behalf; competence is the idea that people feel appropriately matched to their given life and work tasks, and are thus able to get wanted results; and relationships with others are the presence of relationships that include respect, trust and caring between people (Ryan and Deci 2002; Deci and Ryan 2008). The idea is that through fulfilling those basic intrinsic needs, people activate their inner development, are increasingly reliable, enlarge their mental and physical well-being spheres, and are more in line with their true selves (Deci and Ryan 2006). Moreover, it supports the acceptance and internalization of external principles and goals, which eventually leads to more motivation, productivity and a greater willingness to perform and help (Mende, Bolton, and Bitner 2013). However, these basic intrinsic needs cannot be satisfied by individuals themselves which is the pro-social aspect of SDT. All human beings need a certain amount of autonomy or certain kind of relationship with others in order to increase their well-being, but they cannot influence the fulfillment of those criteria, as the criteria are external (i.e., in order to have relationships, one must have friends). Individuals should then work in tandem to increase well-being of themselves, thereby increasing well-being overall.

Human Flourishing

Individually and separately, hedonic and eudemonic well-being research have dominated the positive psychology field (Diener and Seligman 2002; Seligman and Csikszentmihalyi 2000; Deci and Ryan 2006), but the major authors have yet to found a compromise between them. Human Flourishing (HF) was presented by Cambridge University scholars Felicia Huppert and Timothy So as "a combination of feeling good and functioning effectively" (Huppert and So 2013) where "feeling" is a synonym for the hedonic and "functioning" for the eudemonic aspects of well-being. Their approach defines HF as the mirror opposite of widespread mental illnesses. Further, they are defined in a way that allows for denomination of their mirror opposites. A panel of three experts and one lay person developed each item as the mirror opposite of a symptom of the mental disorders depression or anxiety. They continued their study by identifying questions from the rotation module "Personal and social well-being (section E)" of the European Social Survey (ESS) 2006 (Jowell et al. 2006) that are best suited to cover the said items. One question was selected per construct, with such items that have a long-term connotation in favor of short-termed ones. The resulting questions and associated items are presented in Appendix I.

By testing for the distribution of the respective scores per item in the general population (based on the ESS dataset), and their correlations, Huppert and So developed an operational definition by calculating pe is the single item "positive emotion", c_j as the items of "positive characteristics", and f_k those of "positive functioning"; where l and m are the respective item counts per group.

$$HF = pe * \left(I_c * I_f * \left(\sum_{j=1}^{l} c_j + \sum_{k=1}^{m} f_k \right) \right)$$

$$I_c = \begin{cases} 1, & if\ |P_c| \geq l-1 \\ 0, & else \end{cases}$$

$$I_f = \begin{cases} 1, & if\ |P_f| \geq m-1 \\ 0, & else \end{cases}$$

$$P_c = \{c_j : c_j > 0\}, \quad P_f = \{f_k : f_k > 0\} \tag{2.3}$$

Results of a structural equation model show that only positive emotion is a construct of hedonic well-being, the other nine measure eudemonic well-being (Huppert and So 2013). This emphasizes the importance to treat positive emotion as a single item whose absence prevents to classify an individual as being flourishing. (Huppert and So 2013) present a middle-ground approach by combining then validating an instrument that considers hedonistic and eudemonic elements of well-being with single-item measurements.

2.2 Discussion: An Interdisciplinary Definition of Well-being

In summary, both hedonism and eudemonism have been proposed as the ground truth of well-being for millennia before being the object of study in the field positive psychology (Ryan and Deci 2001). Until now there is not a singular definition in place. Due to the complexity of defining well-being, there is no right answer on how to measure well-being (Samman 2007; Ahn et al. 2011; Veenhoven 2008). Currently discussed well-being measures either aim to measure participants' instantaneous well-being (SWB) or dimensions amounting to wellness (PWB). Measurement matters: the employed scale dictates if the assessment can be used as a reflection of satisfaction (ex-post) or as a tool of design (ex-ante).

SWB is temporally oriented, focusing on the individual feeling of happiness as calculated by the presence of positive emotion and absence of negative emotion (Diener 1984a; Kahneman and Krueger 2006; Kahneman et al. 2004b). PWB allows for an alternative view of well-being, namely that what feels good and what makes one happy doesn't (always) lead to a meaningful expression of well-being or acting with integrity (Waterman 1993; Waterman 2007). However, in attempting to measure the conditions of well-being and not only the feeling, PWB becomes so hyper-dimensional as to become non-assessable. Specific instruments have been developed for assessing the main determinates of PWB, the most commonly applied thereof being Self-Determination Theory. In measuring individual's perceived self-determination, competence, and relationships with others rather than general subjective assessment, (Deci and Ryan 2008) argue that individuals' summed well-being is correctly estimated. Human Flourishing is introduced as a hybrid of hedonic and eudemonic well-being. The separate measurement systems have failed to take all aspects of well-being into account until now, which makes Human Flourishing especially attractive as a well-being indicator in progressive community management.

Table 2.1: A comparative assessment of psychological instruments of well-being assessment

		Single Item	Multi-Item	Time Series	Momentary	Real-time	Data-Pulled	Data-Pushed	Data Sources
Diener 1984, 1994	**SWB**	○	●	○	●	○	●	○	Medium *n* directed questionnaires
Waterman 1993, 2007	**SWB**	◑	●	○	●	○	●	○	Medium *n* directed questionnaires
Deci & Ryan 2008	**PWB**	●	○	○	●	○	●	○	Medium *n* directed questionnaires
Kahneman 2002	**PWB**	●	○	●	○	●	○	●	Small *n* directed questionnaires
Huppert & So 2011	**HF**	●	○	◑	●	○	●	○	Ex-post national surveys
This Work	**HF**	◑	●	◑	◑	●	◑	●	Small *n* directed questionnaire

○ not covered ◑ partially covered ● covered

Table 2.1 is a comparative view of the major psychological contributions to well-being. It assesses the item measurement (single or multi-item questions), the timespan with which the authors validated their instruments (longitudinal time series, momentary (cross-sectional) assessments, or real-time assessments), and if the data was solicited by the researcher (pulled), or if the data was volunteered by the participant (pushed).

As HF provides a *fil-rouge* between hedonic as well as eudemonic well-being it reduces the risk of what Aristotle saw as the 'slavish pursuit of desire' (Ryan and Deci 2001) embedded in exclusively hedonic approaches. Moreover, the diversification of well-being across positive emotions, functioning, and characteristics reduces the impact of single item measures. Overstatement and misinformation, widely reported in SWB measures, are therefore less likely and less impactful when they do occur (Veenhoven 1984).

Human Flourishing is taken as the operationalized definition of well-being for this thesis (**RQ 1.1**). HF is an elegant solution that simultaneously measures SWB and PWB, as well as highly granular components of well-being. Further, as mentioned above, the risk of inflated or over-reporting are mitigated with Human Flourishing's triangulated approach. This work builds on the principle that both single and multi-item measurements can provide a valid assessment of well-being. In order to follow the standards of best practice and calibrate participants' base-line well-being, the single-item measurements of SWB and PWB are applied as survey items the form of the HF survey of (Huppert and So 2013). This work also applies multi-item measurement in the form of sentiment analysis, (see Chapter 3.2.3) in order to not only address historical or momentary well-being, but real-time well-being. Finally, whilst the survey items are pulled (solicited) data sources, the majority of the data analyzed is pushed (unsolicited) from participants for unobtrusive and less biased measurement and assessment (discussed in Chapter 3.2.3).

Chapter III Related Work

"Value creation through service provision and service exchange relationships at the micro level must be understood in the context of value creation through service provision and service exchange relationships at the macro level. The elements are value, relationships, and networks; the driving force, and thus the nature of value, relationships, and networks, is mutual service provision for mutual wellbeing."

Toward a Transcending Conceptualization of Relationship:
A Service-dominant Logic Perspective, (Vargo 2009)

Service design is transformative when it has a measurable, even optimizing, positive effect on well-being. This is an exciting approach: irrespective of domain, TSR delivery guarantees well-being outcomes like enabled or increasing access, social justice, social capital, agency, and ecological stability (Rosenbaum et al. 2011).Well-being outcomes here refer to both well-being of the individual and the collective (Veenhoven 2013; Samman 2007). TSR's multidimensionality is nicely highlighted in Ostrom et al.,'s 2010 article:

> *"As such, it [TSR] examines aspects such as the social and ecological consequences and benefits of services offerings, increased access to valued services, the disparity in the quality of service to different groups, the design and co-creation of services with consumers that honors both the agency and the values of individuals and communities, the identification of and planning for the impact of services on well-being and the impact of consumers' service experiences on well-being." (Ostrom et al. 2010, 9)*

The conceptual domains of TSR are extensive and well-covered in the foundational conceptual works of (Ostrom et al. 2010b; Anderson et al. 2013; Rosenbaum et al. 2011), including healthcare, finances, and the workplace. However, such the TSR framework brings about the following, non-domain specific questions: Where is the intersection of personal and communal well-being; and, how granular does TSR need to be in order to establish a robust measure? The coming discussion is an extension of (Hall et al. 2014), where these aspects were discussed in order to ground the discussion of well-being measurement in service dominant logic.

3.1 Service Design for Consumer Well-being

TSR was borne out of the recognition of the importance of services to both the global economy and individuals' daily life; this interplay becomes especially important considering that by 2050 is it estimated that the world's population will approach nine billion.[4] This requires a service-level commitment to human development and quality of life standards from the state, and a convincing statement of managerial necessity and delivery from the private sector: a so-called triple bottom line approach of people, planet and profit (Norman and MacDonald 2004). Service design has a fundamental role in developing this approach by taking both provider commitments and consumer well-being outcomes into consideration (Rosenbaum et al. 2011), thus creating service design that enables well-being.

Transformative service research (TSR), a recently-envisioned branch of service science, is about understanding connections between service offerings and well-being. It has at the core of its conceptualization the goal of improving the well-being of individuals. A founding statement characterizes TSR as: "the integration of consumer and service research that centers on creating uplifting changes and improvements in the well-being of consumer entities: individuals (consumers and employees), communities and the ecosystem" (Anderson et al. 2013). It is clear that in the modern economy, service touches innumerable aspects of daily life. It is then natural that the field of service science explores mitigation of negative and enhancement of positive service experiences beyond the value co-creation and consumer satisfaction paradigms. This is well summed up in the conversation between the switch from goods-dominant to service-dominant logic (Vargo and Lusch 2008; Vargo, Maglio, and Akaka 2008; Vargo and Akaka 2009).

Currently the TSR agenda is lacking a measurement tool that considers the foundational structure of how well- and ill-being implant itself into service-oriented society. In order to use well-being as a societal indicator, that indicator must first be delineated. Mapping well-being, or its negatively correlated partner ill-being, is not such an imminently achievable task. Well-being is per definition highly subjective, multi-dimensional, dynamic, and at best fuzzily defined. As noted by White and Pettit it is important to recognize that the concept under discussion is normative – that well-being and its assessment are inevitably based on values and judgment. This well-being is attributed to states – 'being' in terms of material endowments, psychological attributes, and subjective assessments of the personal and environment one exists in (White and Pettit 2004).

In order to move the TSR agenda forward, an extension to the existing framework of (Anderson et al. 2013) which captures the intersection between service and well-being of individuals, communities, and the ecosystem is necessary. A detailed framework proposal follows in the coming sections.

[4] http://www.census.gov/popclock/. Last Accessed: 12 June 2014.

3.2 A Transformative Service Framework

This framework extension utilizes a systems approach, meaning the entirety of the service environment is considered in order to assure success (Spohrer and Maglio 2010). In addition to Anderson et al.'s macro-level factors it adds meso- and micro-level environmental factors. These aspects (service influencers) are generally considered external to service design, where a service influencer is defined as a cycle of provision, perception, and impact, and well-being outcomes (Figure 3.1). This layered approach allows for analysis of the granularity of daily life; by extending the model with these dimensions, researchers are able to suitably analyze the often compounded aspects of ill-being.

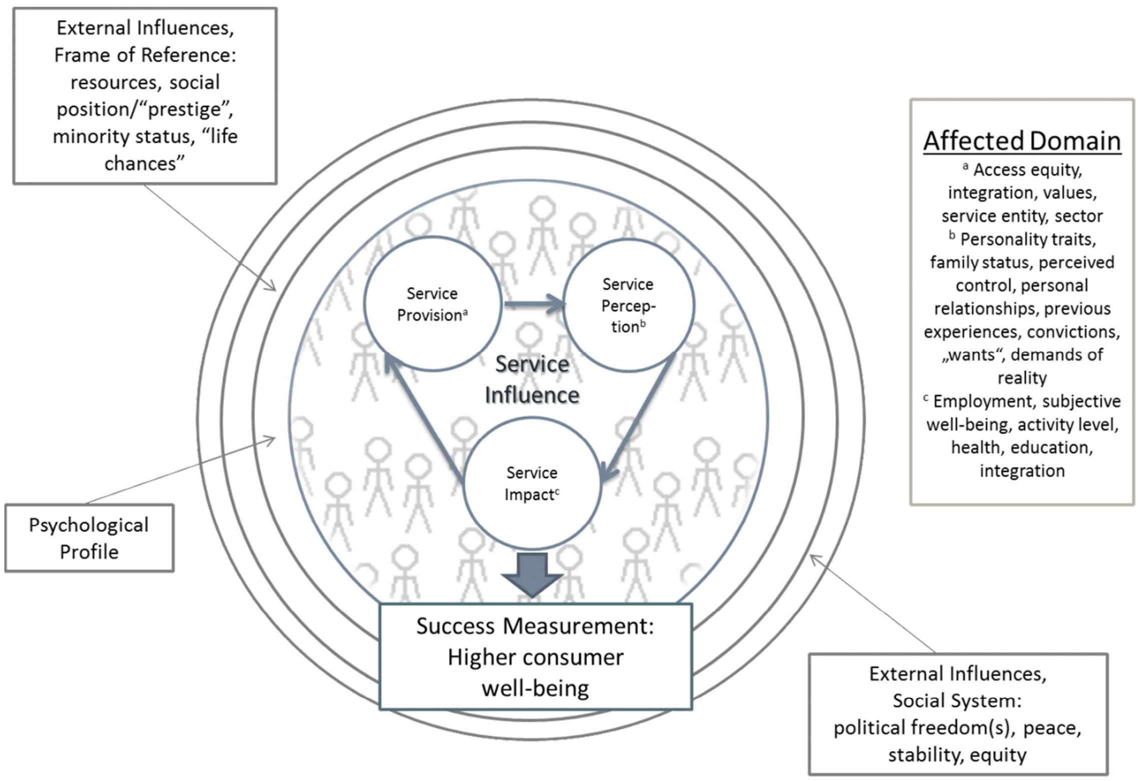


Figure 3.1: An adaptation of (Anderson et al. 2013)'s TSR framework

A fundamental reference point for personal and collective assessment of well-being lies in the greater social system (Stiglitz et al. 2010; White and Pettit 2004). This then must include macro-level assessments like access to political freedoms, general peace and stability, equity and overall development (Anand and Sen 1994) and the meso-level of external frame of reference; i.e., how one perceives their place in society (White & Pettit 2004). Here one finds objective measurements like social hierarchy and minority status, as well as less standard measures like 'life chances' one has had, and the general prestige of their life circumstances (Veenhoven 1984; Veenhoven 2013). In this framework, the micro-level of consumer-service interaction is the psychological profile of the individual. It is well-established that one's baseline psychological profile affects the way one subjectively understands their circumstances overall (Schwartz et al. 2002; Purvis et al. 2011; Hall et al. 2013).

The affected domains referenced in Figure 3.1 have a strong correspondence with macro-, meso-, and micro- environmental factors. Things to consider in transformative service provision include access equity, integration, values, service entity, sector and overall inclusiveness (Anderson et al. 2013; Gebauer and Reynoso 2013). Perception of service provision is driven by a combination of individual and collective understanding of personality traits, family status, perceived control, personal relationships, previous experiences, convictions, and general "wants" balanced by the demands of reality (Veenhoven 1984). The optimal impact domains are those such as employment, SWB, activity level, health, education, and integration (Rosenbaum et al. 2011; Anderson et al. 2013). When TSR incorporates these aspects, the resulting effect should be an increased consumer well-being.

3.2.1 The Outer Circle: Macro-level Influences on Well-being

Within a secure, participatory democracy and a strong economy there are fewer chances for wide disparity levels between subgroups. This implies that each member of society has access, or a reasonable expectation to be able to participate, affording minorities and other subpopulations the chance of equal servicing. This is generally not true for opaque or authoritarian systems: such governments are less likely to be stable and more likely to provision services along partisan, ethnic or religious lines. Not only are groups unequally serviced, but quality of life overall drops with respect to expected welfare maintenance (Wu and You 2007; Lacey et al. 2008; Ballas 2013; Diener and Suh 1997). Changes in the overall well-being of the state are driven from the aggregate number of citizens in the state and their access to (civil) services, reflecting the view that progress is contingent to the impacts on and richness of the human life, rather than merely economic advances (Stiglitz et al. 2010; Buchanan 2001). This is tantamount to the economic, or ex-post, assessment of well-being.

A useful model for the utilization of macro assessment of well-being as a decision making aid was proposed by sociologist James Davies in his 1962 article on social unrest (Davies 1962). He suggests that drops in expectations as compared to actual progress fuels relative deprivation, the idea that deprivation is only experienced when compared to others who are more fortunate (see Figure 3.2). In his model, a significant difference between actual and expected advancement reveals the overall well-being and vigor of the institution. In other words, social unrest is a subjective response to a sudden reversal in fortunes after a long period of growth (Davies 1962). The strength of relative deprivation is evaluated by charting and changing the expected change of actual well-being levels against expected well-being figures. For a given construct of well-being (cf. the discussion in Chapter 2), a lack of statistically significant differences between expected and actual well-being levels implies no discrepancy and no social unrest; significant differences implies the opposite. This is a key research concept: as the model suggests, if relative deprivation is not in effect, social turmoil does not occur regardless of the actual state of well-being. Given a satisfactory answer to Research Question 2.1, this model is employable in the evaluation of Research Question 3.

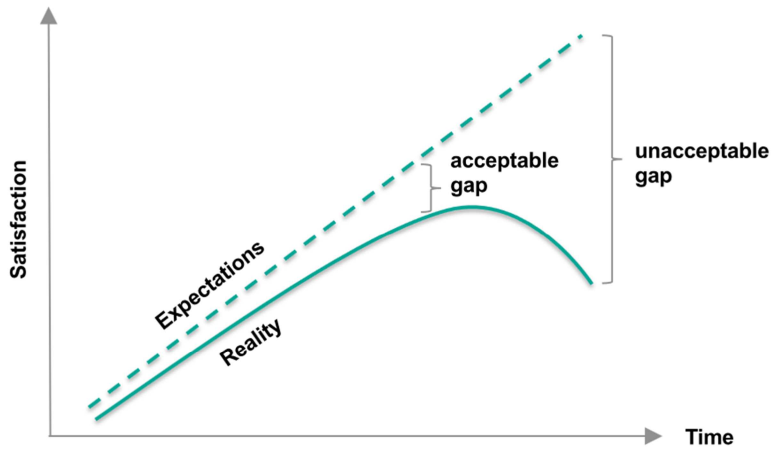


Figure 3.2: The Davies J curve

In his dissertation "Conditions of Happiness" noted Dutch psychologist Ruut Veenhoven wrote "The more healthy and active the citizens and the smoother their contacts, the greater the chance that society flourishes. Moreover, widespread dissatisfaction with life tends to act as a bomb under the social system (Veenhoven 1984, 404)." This is in agreement with the argument of Davies that significant issues of well-being manifest in (sub)groups of the population, and negative well-being will follow a Davies J-curve distribution (Davies 1962). This model indicates when social expectations have a large deviation from the actual outcomes of human well-being (relative deprivation), some form of social schism should be expected (Figure 3.2). A fitting and correct measurement of well-being can be leveraged to provide actual and expected trending of flourishing. With concurrent supervision, components that can cause agency loss (in this case, statistically significant drops in well-being data) can be proactively regulated as a form of adaptive community management. Applications for this sort of management tool are manifold: business, civil society, and public policy can benefit to name a few domains. Such a model has diagnostic value and can be exploited to have predictive worth. The predictive worth of the model is the potential to be used in charting future public participation- based unrest and movements. More concretely, given the community's overarching well-being trends events causing communal spikes and dips in well-being can be pin-pointed and assessed.

3.2.2 Meso-level Analysis: The Role of the Self in the Community

As noted in (Ozanne and Anderson 2010), individuals, structural issues, and the socioeconomic context of a given area must be taken into consideration when completing impact assessments. Knowledge of the preexisting conditions and self-assessed roles of a given consumer group is necessary when designing and implementing services to increase communal well-being and/or decrease communal ill-being (Saatcioglu and Ozanne 2013). Well-being is not only access and psychological health, but the perception of one's place within the greater environment. Individual well-being is intrinsically linked to the individual's perception of belonging in a community, and their relative status within it. These singular assessments aggregate up to communal well-being. This is to say, in areas where high individual well-being exists, there tends to be high communal well-being. In areas of compounded disadvantage, well-being and its related outcomes tends to be low. This is confirmed in the Framingham Heart Study: high and low well-being networks tend to be clustered within three degrees of separation from one another (J. Fowler and Christakis 2008). This is especially relevant for mapping the contours of a community based on its sentiment (Research Question 3).

The proposed meso-level environment for transformative services is closely aligned to George Vaillant's finding on the antecedents of flourishing from the Harvard Grant Study, to date the longest running longitudinal sociological study. He writes that formative experiences are crucial to future health and happiness; the presence of positive relationships matter for happiness; the risks one takes with their lives (e.g. drug and alcohol consumption) have high prediction abilities on one's ability to maintain family and social relationships (Vaillant 2008). Meso-level analysis is not foreign to the TSR agenda: quoting (Ostrom et al. 2010, 9), TSR considers "[…]the disparity in the quality of service offerings to different groups, the design and co-creation of services with consumers that honors both the agency and the cultural values of individuals and communities, […]", which requires an understanding of the person and their understanding of belongingness in their community. Longitudinal surveys, panels, and various forms of network analysis can establish the indicators of the meso-level.

Data gained from international databases and surveys are well utilized at this level. Considering this, and the other well-being oriented indicators from the largest public surveys, and how to parse the various important domains into a taxonomy is an important, ongoing challenge for TSR. Synopses of the largest international and national instruments are below, and a comparison table can be found in Table 3.1.

Kingdom of Bhutan

The Kingdom of Bhutan provides a point of reference of how well-being can be used as a framework for wider stakeholder accountabilities (Thinley 2011; Bhutan 2012). In the late 1980's, the kingdom conjoined externally imposed indicators such as Gross Domestic Product (GDP) per capita and the state of the environment as a measurement of the state of health with a focus on national well-being assessments as the central key performance indicator in its Five Year Plan of development. As stated in the national planning guidelines: "Apart from the obvious objectives of development: to increase GDP on a national level and incomes at the household level, development in Bhutan includes the achievement of less quantifiable objectives. These include ensuring the emotional well-being of the population, the preservation of Bhutan's cultural heritage and its rich and varied natural resources (Bhutan 1991, 1:6)." This statement is clearly indicative of the full inclusion of macro, meso and micro indicators of TSR.

This process has been furthered in two ways: time-lapsed surveys, and well-being framework integration. The surveys give status reports on the health and vigor of the nation, where framework integration serves to further the stated policies of governmental planning commissions. Frameworks of well-being and its conditions are being integrated into public programming and services, as well as national universities and the public bureaucracy (Bhutan 2012). Impressive results ensued: According to the United Nations Development Programme since the inception of its well-being focused Five Year Plans, Bhutan has made major strides (Kumar et al. 2007). Its GNI per capita of $1,005 (in 2005 dollars) was 40% higher than that of India, and over 70% higher than the average income of low income countries. The country's human development index grew from 0.325 in 1984 to 0.583 in 2003, placing Bhutan in the category of medium human development countries (Kumar et al. 2007). In implementing an enhanced indicator series Bhutan has a more reactive, finer tuned, and richer set of data from which to base its policy decisions.

European Union

There has been an upwelling of attention directed at understanding and measuring well-being as a conceptual and practical compliment to myriad macro and micro indicators and as policy and decision making tools. A prominent example is the Commission on the Measurement of Economic Performance and Social Progress, formed by Nicholas Sarkozy during his term as president of France (Stiglitz, Sen, and Fitoussi 2009). This working group and report are the most notable examples of reconfiguring "standard" measurements and related constructs as measures of national progress and well-being. This study concentrates mainly on the macro and meso indicators of the TSR framework. Due in part to its provocative findings, on-going efforts are in place across the European Union and worldwide.

The United Kingdom's Office of National Statistics is most comparable to the TSR framework in the European Union. It publishes overview data of national well-being twice a year, in addition to a European comparison report. The reports take care to highlight particular communities of interest; children, minorities and recent immigrant to name a few. This reporting series is notable as it, like Bhutan, integrates national, communal, and personal well-being indicators in its assessment. It is also the most fully integrated system of well-being assessments at the national level in the European Union. Not only policy makers but the public has access to review and comment on the drivers of well-being in the United Kingdom due to their open statistics API.

In a similar effort, the German federal government conducted a national study called "Growth, prosperity, quality of life - Towards a sustainable economy and social progress in the social market economy" in 2013 (van Suntum 2012). They argued that GDP is no longer a complete picture of the quality of life in Germany, and the German people and the government need a more complete overview of the quality of life of the Germans. An "improvement of statistics is necessary [...]" (van Suntum 2012), and policy goals based on better assessment of what makes a happy, health community is a contemporary solution to this challenge (Ballas 2013). Thereby the German Parliament proposed ten new criteria to measure the country's health and wealth. The most significant additions from the perspective of progressive policy making of the new criteria are the indicators material well-being, social affairs and societal inclusion (all meso indicators of TSR), as well as ecology (a macro indicator of TSR).

Eurobarometer

The Eurobarometer survey[5] is taken twice yearly at the behest of the European Commission's Directorate-General for Communication and is aimed at gauging public opinion in (and largely about) the European Union. Its focus is not on happiness or well-being per se; rather, it aims to assess public attitudes (in all 27 members of the European Union) towards matters of public import in the EU. In the context of TSR, this is a complement to surveys such as the General Social Survey (GSS) that aim to measure well-being directly. The Eurobarometer series measures PWB of the individuals associated with, and affected by the EU. For the purpose of TSR, the EU exemplifies a service-providing institution and the Eurobarometer survey illustrates how one such institution measures its performance in the eyes of its clients. It is worth noting that the EU, as of the last available report,[6] is in turmoil due to continuing effects of the major worldwide economic recession of 2008, including the continuing financial crises of Greece and other EU members, and the continuing struggles with other major policy decisions. For present purposes this makes the EU a highly interesting institution. How do the EU's well-being assessments (broadly construed) reflect this turmoil?

While the absolute levels of prevalence of various opinions are surely important, arguably, changes over time are at least as valuable for policy design and institutional assessment. Significantly, the Eurobarometer report emphasizes throughout the dynamics of the attitudes it reports. The attitudinal variation among the 27 EU members is often strikingly large. In the spring of 2012 the survey found that those giving their country and overall "good" assessments ranged from 83% in such countries as Sweden, Luxembourg, Germany, and Finland to 0% in such countries as Greece, Spain, Portugal, and Ireland. This range narrowed in the fall 2012 survey from 75% to 1%. This is hardly an improvement, although it is consistent with the

[5] For more information see http://ec.europa.eu/public_opinion/index_en.htm. Last Accessed: 17 June 2013.
[6] This is available at http://ec.europa.eu/public_opinion/archives/eb/eb78/eb78_en.htm. Last Accessed: 18 June 2013.

finding announced in the report that attitudes have been roughly stable of late. Looking at the EU, member states constitute a natural categorization by which to measure attitudes. But there are other natural categorizations as well which need to be considered, for example, by age, gender, occupation, and income. Even more so, people are multi-dimensional, which means that they will fall into several categorizations at once. What are the particularly vulnerable profiles? The larger meaning for TSR and for measuring well-being in smaller-sized institutions is that attitudinal variation may be critically conditioned on categories that may or may not be identified. Recognizing these categories should be seen as a continuing challenge for TSR.

OECD Better Life Initiative

The Organisation for Economic Cooperation and Development (OECD) collects statistics and survey data extensively. Most relevant to TSR is the OECD Better Life Index.[7] The OECD's Better Life Index[8] is composed of 11 "topics" (measured either by a single indicator variable or by an index of a small number of indicators). These meso indicators are: housing, income, jobs, community, education, environment, civic engagement, health, life satisfaction, safety, and work-life balance. The data for the Better Life Index also supports a degree of online analysis, and is fully comparative. In addition, links are available to the very large number of other data collections created and maintained by the OECD. Many of these will also be of interest to TSR scholars for the breadth of aspects which are covered.

International Social Survey Program and the General Social Surveys

The International Social Survey Programme (ISSP), at http://www.issp.org/, is the international umbrella organization coordinating the GSS management and archival of 48 countries. These countries are predominately developed countries, although some interesting statistics are available, such as those from China and Venezuela (two otherwise opaque countries). The ISSP and GSS have maintained the major of their questions since the inception of the survey in order to facilitate and longitudinal and replication of the information. The 1972-2012 GSS has 5,545 variables, time-trends for 2,072 variables, and 268 trends having 20+ data points.[9]

The GSS waves contain a standard 'core' of demographic, behavioral, and attitudinal questions, plus topics of special interest specific to a given wave. The GSS data are downloadable in various formats friendly for statistical processing. The website also makes available a basic online analytics capability for the data. The GSS specializes in trend data. Especially

[7] http://www.oecd.org/statistics/datalab/bli.htm. Last Accessed: 7 March 2015.
[8] Accessible at http://www.oecdbetterlifeindex.org/about/better-life-initiative/; the data used to create the index may be found at http://stats.oecd.org/Index.aspx?DataSetCode=BLI. Last Accessed: 7 March 2015.
[9] Available at: http://www.issp.org/page.php?pageId=4. Last Accessed: 12 June 2013.

distinguishing in comparison with the other collections discussed, the GSS site lists about 300 published articles that use its data. The GSS is high quality, broadly scoped source of survey data pertinent to TSR. Of all the sources reviewed here, it is likely the one that has been used the most in scientific publications.

Table 3.1: National and international well-being measurement instruments

		SWB	PWB	Economic Well-being	Institutional Integration Level	> Yearly	Yearly	< Yearly	Data Sources
Bhutan	National	●	◑	●	●	○	●	○	Interviews & Questionnaires
France	National	◑	●	●	○	○	○	●	Ex-post Indicators
Germany	National	●	◑	●	○	○	○	●	Ex-post Indicators
United Kingdom	National	●	●	●	●	●	○	○	Questionnaires
Euro-barometer	International	●	●	●	◑	●	○	○	Interviews
Better Life Initiative	International	◑	●	●	◑	○	●	○	Questionnaires
General Social Survey	International	●	●	◑	◑	○	●	○	Questionnaires

○ not covered ◑ partially covered ● covered

As seen in Table 3.1, the most complete well-being instrument is located in the United Kingdom; it is however limited to Britain, Scotland, and Northern Ireland. The Eurobarometer is much more expansive, though its institutional integration is limited at making suggestions for increasing well-being of European citizens. It has the further limitation of being interview-based, indicating that only small proportions of the citizenship can be addressed at any point. Both France and Germany currently concentrate on ex-post macro indicators; while a laudable start, such indicators can no longer be understood as a proxy for well-being due to their macro nature, the time-lagged delay in data collection, and too-broad definition (as discussed in Chapter 1). It can be seen that while data is being collected at the national and international level, still be the implemented is a well-being indicator feeding into a TSR application that is near to real time, with low-cost and scalable data collection methods.

3.2.3 Me, Myself and I: Micro Profiles and Well-being

As mentioned earlier, an important factor in well-being is the baseline psychological profile of the person. Considering psychological profile is of upmost importance when measuring service perception as shown in Figure 3.1, as it is well-established that different personality types report satisfaction and well-being with difference reference points. Confirmed in multiple studies, psychological factors like low(er) needs for circumstance maximization, psychological needs satisfaction, personal goal progress, high self-esteem, and a positive Big Five Inventory[10] profile are prerequisites for high well-being (John et al. 1991; (B. Schwartz et al. 2002; Purvis, Howell, and Iyer 2011; Hall, Caton, and Weinhardt 2013; John, Donahue, and Kentle 1991; Sheldon and Hoon 2013).

Maximization refers to one's ability to be happy with a decision once it has been met. The more one "maximizes" a decision making scenario, the less happy one is in the long term, 'the paradox of choice' (B. Schwartz et al. 2002). Considering psychological needs satisfaction, (Sheldon and Hoon 2013) modeled optimal human well-being with a hierarchical regression analysis, finding that there are four tiers of personality which are predictors of well-being. Their work shows that social relations, self-narratives, goals and life intention, personality traits, and psychological needs are all necessary for high well-being. The Big Five personality factors is the most well-known and widest used personality traits model in psychology, human resources, and a plethora of other institutions (John, Donahue, and Kentle 1991). A well-being inducing or positive Big Five profile is considered to be low neuroticism, high extraversion, and a combination of optimism, agreeableness, conscientiousness in the terms of this thesis (Purvis, Howell, and Iyer 2011; Hall et al. 2013; Sheldon and Hoon 2013).

This level presents the most problematic measurement area. Institutionally defined and managed well-being requires a high level of trust between participants and stakeholders; the design of transformative services requires substantial participant support and participation. Generally speaking, psychometrics are left for the domain of psychology and are strictly outside of service design and policy-making. This is because the type of data could be used to observe not only public but also private life domains. Whereas responsible designers use well-being to view the institution's overall progress, satisfaction, and capacity, irresponsible management could use well-being data to pin-point those who do not "fit in" with institutional standards or desires, as well as the risk of identification of reportedly anonymous participants (Zimmer 2010). Other irresponsible uses of data can include harm by incidentally altering the well-being of (unwitting) participants as was seen in the study on emotional contagion by (Kramer, Guillory, and Hancock 2014). This is especially relevant in the case of participants with a high vulnerability level as assessed by the meso-level interaction (Markham and

[10] The Big Five are Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

Buchanan 2012). Participants will need to place significant trust in stakeholders to ensure validity and reliability of the data (such as the example in the United Kingdom).

With potential issues recognized, the prospective uses for psychological factors to TSR are still manifold. Research designs for establishing this level include ethnographies (e.g., (Saatcioglu and Ozanne 2013)) and psychometric surveys (Kahneman et al. 2004b; John, Donahue, and Kentle 1991). Both methods are considered expensive in terms of funds and time. Therefore, researchers are concentrating on less expensive mechanisms to measure psychometrics, especially considering the digitalization of daily life since the advent of the internet. The coming sections introduce state of the art mechanisms for the measurement of well-being.

On the Application of Social Media Platforms for Social Sentiment Analyses

"Social Media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user generated content" (Kaplan and Haenlein 2010, 61).

Social media enables researchers to collect and analyze large scale, unobtrusively gathered, individual data. Researchers previously faced two common obstacles. Firstly, even if social data is gathered at a sufficient level, information is often spread over various agencies, precluding efficient analyzing processes. Secondly, it requires steady collection updates over time to register movements in social characteristics (Hackenberg 1970). The longer the time span between updates is, the less accurate the data and thus the analysis can be, as several other reasons might have occurred in the same time interval.

In the late 1960s computational innovations resulted in a shift of challenges: The restricting parameter for work of social researchers was no longer the processing of data. Instead, information grew at a rate faster than researchers could analyze (Cioffi-Revilla 2010). Considering the decades since the beginning of globalization, quickly developing (digital) technology and fast moving economies, the developments in people's daily lives become at once more transparent, yet more difficult to understand. This is due in part to the rise of networked, social data. Hand in hand with technological and digital evolution is the capability to collect and process information. Modern social data shares these attributes:

1) Large (easily) extractable amounts of data

2) Continuous data streams over time

3) Spatial and design independence for researchers

Social media sites in particular have quickly ascended from a novelty of the early 2000's to a fact of life, and daily necessity. Today, Facebook is accessed daily by ten times more people than the population of Germany.[11] Users interact online by creating profiles and providing (semi)personal information in form of text, photos and other media (Röll, 2010). Röll summarized that while motives for using the Social Networking Sites range from staying in touch with fellow friends and dating services to establishing professional business networks, all pages share predefined rules how social connections are made. These rules are what determine the resulting social network. In most cases mutual acceptance is required to link two profiles (e.g. Facebook and LinkedIn). Exceptions exist: On Twitter and Google Plus (to some extent) any user can receive information from any profile of interest. These connections define how users can share and receive different kinds of user generated content.

Due to the fact that social networking and media platforms are generally based on true identities or variants thereof (Lingel, Naaman, and boyd 2014), they are well suited to display online communities. Facebook is the largest platform and with its 864 million daily active users in the end of 2014 (1.35 billion monthly active users) is also the most active one, with one in every seven minutes worldwide (and for Americans, one in every five minutes) being spent on Facebook.[12] Facebook requires mutual agreement for users to link as friends. User generated content can be shared via posts which appear on 'timelines' of users, pages and groups. Users may further share content by referring to an already existing post with a commenting function. Users control privacy by defining rules for individuals or groups, and private or targeted messages are allowed, assuming the recipients' privacy settings allow for it. Facebook offers the feature of 'Pages' that differ from the standard user profiles. Unless specifically restricted in the page's settings, the information on these pages are completely public.[13] This important distinction from user profiles allows researchers to gather data of most publicly acting online communities without further requirements.

In an exhaustive survey, (R. E. Wilson, Gosling, and Graham 2012) summarized and classified 412 articles written on Facebook for the period 2007-2012 leading to five supra-categories: descriptive analysis of users, motivations for using Facebook, identity presentation, the role of Facebook in social interactions, and privacy and information disclosure. The review addresses key articles across these five categories, and the methods employed by the various scholars. Recognizable is that the usage of Facebook's API by non-Facebook staff or partners to support unobtrusive studies is low; when the referenced studies apply quantitative methods, the method of choice tends to be based in survey methods.

Notable studies from Facebook Research look at public expressions of sentiment. (Kramer 2010) used status updates based in the United States to create a composite well-being index.

[11] http://newsroom.fb.com/company-info/. Last Accessed: 12 March 2015.
[12] http://techcrunch.com/2014/07/23/facebook-usage-time/. Last Accessed: 12 March 2015.
[13] https://www.facebook.com/help/387958507939236. Last Accessed: 7 May 2015.

This has since been criticized in (Wang et al. 2014), who state that Facebook status messages are not appropriate for well-being assessment, but rather mood regulation. Another series of studies by Kramer and colleagues (Kramer 2012; Kramer, Guillory, and Hancock 2014) reviews emotional contagion on Facebook. These studies report that emotions are indeed contagious in a network. Their findings support that short informal text like Facebook status updates can be used to measure sentiment online. Further confirmation can be found in (H. A. Schwartz et al. 2013), who collected and analyzed 74,941 Facebook profiles with LIWC and were able to establish linguistic characteristics of personality, gender, and age. In depth discussions on the use of Facebook in sentiment analysis can be found in Chapters 5.1 and 6.1.

Gamification as an Incentive Mechanism

In gamifying well-being, leaders take proactive steps towards smart community management. Acting as a thermometer by which to gauge institutional health, well-being data serves not only as a feedback mechanism between various actors and policy makers, but as a forward-looking decision making tool (Ahn et al. 2011; Frey and Stutzer 2007). Thus there is wide-spread interest in tracking mechanisms with high popular acceptance. Until recently, attempts to collect well-being data as an institutional feedback mechanism have been scarce. More recently, a number of other platforms exist that bind some or all of the principles of online social networks, well-being, and gamification. Some of the most popular and notable examples include Superbetter,[14] the Wellbeing Game,[15] the Happiness Initiative,[16] and Track Your Happiness,[17] though this list is by no means a comprehensive list of all well-being and happiness measurements available online. Such platforms either attempt to increase personal well-being and happiness via tips and tricks (Superbetter, The Happiness Project, the Wellbeing Game), perform basic measurements and trends of happiness reporting (Track Your Happiness), or are a hybrid of both (the Happiness Initiative). Of particular interest are platforms which elicit well-being reports, as they functionally serve as a stated preference data collection method with respect to happiness and well-being.

Emerging work from Vella and Johnson is especially valuable in clarifying the use of gamification in terms of Human Flourishing (Vella and Johnson 2012; Vella, Johnson, and Hides 2013). Their work matches each of the ten Human Flourishing items with up to date findings from the gaming literature. Focusing on studies which relate to well-being or mental health of gamers, this work neatly ties the two sometimes disparate worlds of happiness research, gaming, and collaborative computing. This work does not however propose the design or mechanisms for a well-being game. One idea is the use of social networks, as they can be extended by platform features if a gamified application is designed for use within a

[14] https://www.superbetter.com. Last Accessed: 18 December 2013.
[15] www.thewellbeinggame.org.nz. Last Accessed: 18 December 2013.
[16] http://www.happycounts.org. Last Accessed: 18 December 2013.
[17] http://www.trackyourhappiness.org. Last Accessed: 18 December 2013.

social network (Hall et al. 2012). Besides the social features “leaderboard” (social comparison) and “sweepstakes”, social sharing (“gifting”) gains importance. The incentives “bragging” (notification of one’s social network of achievements) and “inviting” (advertise usage within one’s social) extend the toolbox of gamification methods and serve at the same time as a spreading mechanism for the gamified application (Siegel 2012).

Despite earlier use, the term “gamification” did not see widespread adoption before 2010 (Deterding et al. 2011). Since then different parties have used it with different scopes and connotations. An often-cited definition is that of Deterding. It tries to incorporate the different viewpoints and areas of applications by generically subsuming: “Gamification is the use of game design elements in non-game contexts” (Deterding 2011, 9). However, not all agree. Based on their background in service marketing, Huotari and Hamari, for example, state that it depends on the individual perception of a user if a service is gameful, making it impossible for a service designer to identify the non-game context central to Deterding’s definition (Huotari and Hamari 2012). They specify gamification as “a process of enhancing a service with affordances for gameful experiences in order to support user's overall value creation” (Huotari and Hamari 2012, 19) – prioritizing the of creating better experiences instead of achieving them. The current discussion also covers the transformational opportunities brought through gamification, namely the positive effects that gamification can foster in crowdsourcing or in collaboratively changing the world for the better (Stampfl 2012).

The next discussion point becomes applying game design elements in an effective way. A commonly shared and expressed finding is the separation of human motivation into intrinsic and extrinsic components, with current gamification approaches largely (only) supporting the latter one. Siegel therefore suggests taking special care to create a plausible, linked, and in difficulty increasing system of leveling in gamified applications (Siegel 2012). “Leveling” refers to the progress a user makes in discovering the possibilities of an application. He states that ideally several pathways, tailored to varying personal interests, should guide the user in exploring more comprehensive features. Antin and Churchill argue that motivation and social engagement are not automatically supported by using badges: They posit a dependency from the activities that badges are to award and from context. They discern the five functions – goal setting, instruction, reputation, status/affirmation, and group identification – stating that “the fun and interest of goal seeking is often the primary reward itself” (Antin and Churchill 2011, 2) and that the (wrong) usage of badges could even reduce a user’s intrinsic motivation.

The possible reduction of intrinsic motivation by deploying extrinsic motivators is also described by Deterding who hints on the dependence from social situation or context. He argues that supporting a leaderboard with cash incentives counters a user’s autonomy and thereby intrinsic motivation (Deterding 2011). Further context sensitivity is brought in by Dixon who presents several models for Player Types – each with differing core motivations for playing – and who states that gender and age are an influence to playing motivations and behavior (Dixon 2011). A possible solution besides “personalizing” the respective system through detection of a user’s personal type, Vassileva suggests letting the users choose their preferred goals within the gamified application according to prior intrinsic motivation. This can include showing different (or “exaggerated”) data according to the choice. The common separation of human motivation by intrinsic and extrinsic components is extended by a social one. Two elsewhere in literature not often seen incentives are illustrated: social comparison and community collaboration and quests as a form of challenge that can be resolved by cooperation amongst users, occasionally including time limits (Vassileva 2012).

Gamification is a quite obtrusive method of eliciting data, in addition to the fact that any data obtained in this process is per definition stated preferences (estimates of behavior) rather than revealed preferences (actual behavior). Both aspects have a place in behavioral modelling. In order to address revealed behaviors, another method is required. This is discussed below.

Text and Sentiment Analysis Tools

In terms of a revealed model, text and sentiment analysis is a promising mechanism. Text pulled from social media has the benefit that it is largely unspoiled by research design, and offers a highly granular view of the posting individual. Using short informal text as the foundation of public sentiment measurement differs from other text due to the shortness of the text and the different language used (Thelwall et al. 2010). Word count restrictions, the usage of abbreviations and emotional tokens is fostered, leading to informal text containing slang, abbreviations, and emoticons in various forms and styles as well as truncated sentences (Wang et al. 2014). While this type of short informal text challenges Natural Language Processing, the existence of items like emoticons can help to understand the intended sentiment. Emotive values can be established by human readers or automated text analytics programs. Human-centric approaches have a long history and are well applied in varied domains (Hsieh and Shannon 2005; Kassarjian 1977), but lack scalability. When dealing with the volume required by Big Data analyses, either crowdwork (e.g., (Hall and Caton 2014; Paolacci, Chandler, and Ipeirotis 2010)) or automated programs (Balahur and Hermida 2012; Kim et al. 2006) are generally required. Crowdwork for the analysis of items like status updates and tweets however posed both ethical issues (Markham and Buchanan 2012), and can run afoul of the platforms’ terms and conditions. Two mechanisms are widely used to support the automated recognition of written sentiment: corpus-based approaches and dictionary-based approaches (Turney and Pantel 2010). The corpus-based approach is based on the co-occurrence of words and relies on the latent relation hypothesis, stating that words with similar meaning or sentiment co-occur more often in a sentence or passage than words expressing differing sentiment (Turney and Pantel 2010). Given a core set of known and evaluated words, this methodology identifies words with similar orientation. This approach can be especially useful when trying to search for instances of sarcasm or ironicism which is otherwise lost in the dictionary-based approach (Liu 2010). However, it requires a huge corpus to cover most of the words within the respective language.

Dictionary-based approaches use predefined word lists containing sentiment-loaded words. By scanning the considered text, sums of positive and negative affect can be derived, usually normalized regarding the length of the overall text. Kramer subtracts said sums to get a one-dimensional measure of sentiment (Kramer et al. 2004; Kramer 2010), whereas Golder and Macy argue the independence of both dimensions by measuring them separately (Golder and Macy 2012). While Kramer has used the Text Analysis and Word Count program that was built upon the Linguistic Inquiry and Word Count (LIWC) 2007 dictionary, Golder and Macy directly used utilized the LIWC 2001 dictionary (Tausczik and Pennebaker 2010; Pennebaker et al. 2007). Other dictionaries e.g., SentiWordNet (Baccianella, Esuli, and Sebastiani 2010) or OpinionFinder (T. Wilson et al. 2005) are also available. Whereas SentiWordNet sums up possible positive and negative sentiment and the third term of “neutrality” (Baccianella, Esuli, and Sebastiani 2010), OpinionFinder has its focus on classification of subjectivity and objectivity within sentences (T. Wilson et al. 2005). To date, both lack linguistic localization, a feature making LIWC’s 13 available languages favorable.

The dictionary-based approach, however, is unable to find domain specific orientations and context oriented sentiment (Thelwall et al. 2010). Included in (Dodds et al. 2011) sentiment analysis are tweets surrounding Osama bin Laden’s assassination and the end of the blockbuster show ‘Lost’. It marked May 2, 2011 one of the most negatively affected days within the Anglophone twittersphere due to words like “dead”, “killed” and “terrorist.” Lost’s finale also resulted in a distinctive drop in happiness on the day it was released, but it was not due to sadness over the show ending. The word ‘lost’ was tagged as a negative sentiment word in the utilized dictionary and therefore scored all mentions negatively. Table 3.2 gives a brief overview of the most widely used sentiment analysis packages.

Table 3.2: Comparison of existing dictionary-based sentiment analysis packages

	Corpus-based	Mixed Method	Dictionary-based	Localization	Automated Functioning
WordSmith	●	○	○	●	○
General Inquirer	○	●	○	○	○
Senti-Strength	○	●	○	◑	●
SentiWord-Net	○	○	●	○	●
LIWC	○	○	●	●	◑

○ not covered ◑ partially covered ● covered

In addition, each tool has positive and negative attributes making it more and less suitable for the use of sentiment analysis for Transformative Service Research. These attributes are summarized in Table 3.3.With this consideration set LIWC shows itself to be an especially interesting tool for application in online social media use cases.

Table 3.3: National and international well-being measurement instruments

	Strengths	Criticisms
WordSmith	Context-sensitivity allows for higher accuracy in representing the meaning of the text	1) Corpus-establishment is a complex task and a bad corpus leads to poor results 2) Unproven with Online Social Media data
General Inquirer	Allows for sophisticated context analyses	Complicated adaption processes for different studies restricts number of analyzed categories in practice
Senti-Strength	1) Basic context consideration for booster words to scale emotion (e.g. ‘very’) 2) Specialized for short informal texts (e.g. internet expressions, abbreviations)	Restricted to emotion valence only
SentiWord-Net	1) Robust results for emotional valence detection 2) Extended valence scale (includes ‘objective’ as neutral)	Restricted to emotion valence only
LIWC	1) Flexibility (editable dictionaries) 2) Applied to Online Social Media use cases 3) Easy analysis of broad language dimensions	Missing context observance leads to misinterpretations

Linguistic Inquiry and Word Count

LIWC originally was not intended to be used on short informal text, but to analyze text of expressive and therapeutic writing sessions usually containing more content than the average tweet or Facebook update (Wang et al. 2014; Tausczik and Pennebaker 2010). However, its expansive psychometric dictionary offers a unique opportunity to reveal the latent emotional context of text-based data. LIWC has been shown to possess excellent precision and recall abilities with high but not overfitting correlations in the analysis of latent sentiment (Salas-Zárate et al. 2014; Mahmud 2014), though its performance in prediction tasks is often low compared to n-grams or machine learning approaches (Komisin and Guinn 2012; Balahur and Hermida 2012). The application of LIWC on documents returns the percentage of words across the categories social processes, affective processes, cognitive processes, perceptual processes, biological processes, work and achievement, as well as punctuation and structural details (Pennebaker et al. 2007; Tausczik and Pennebaker 2010). Per cent based information gives the researcher a mechanism by which to see the relative worth of categories in speech. This facilitates measuring change, looking for group-based patterns, monitoring individual spikes and dips, and identifying psycholinguistic profiles.

LIWC's development and validation was an iterative process of word collection, during which several rating scales, standard dictionaries, and experts were consulted. The resulting broad list was coded by three independent judges' who first indicated if a word should or should not be included, then categorized words according to conceptual lists. Their work was then externally validated for psychometric validity in a process that took three years (Pennebaker et al. 2007). Two versions of the LIWC dictionary currently exist -2001 and 2007 – and it is available in 12 languages to date.[18] Several studies have shown its proficiency with short, informal text (Lin and Qiu 2013; C. Chung and Pennebaker 2014). This is not a trivial statement. Social media sites drastically limit word counts of single authors compared to traditional sources (Kramer et al. 2004). Abbreviations (e.g. "howru" for "how are you"), purposely misspelled words (e.g. "helllooo"), special phrases ("lol") and emoticons (e.g. ":)"), which are pervasive in short, informal online texts usually cannot be processed by sentiment analysis toolkits (Wang et al. 2014).

The previous section discussed the importance of contextual settings to avoid misinterpretation of words and complete sentences. In addition, in science exists serious interest in automated content detection of documents, an important branch of text analytics (Lazer et al. 2009; Balahur and Hermida 2012). When people share (written) information, there is not only content but also the way they create their message and the linguistic style (C. Chung and Pennebaker 2007). They found that function words are well suited to build a systematic picture of this inconceivable dimension as latent indicators. They refer to pronouns, prepositions, articles, conjunctions, and auxiliary verbs and altogether can be imagined as "[…] the linguistic "glue" that hold content words together" (Groom and Pennebaker 2002). While LIWC focuses on function words it also includes content words. The functionality is based on dictionaries that assign over 4,500 words to 70 different categories, ranging from a simple stylistic (e.g. article, prepositions) to a complex psychological level (e.g. positive emotion, cognitive words). Due to their near constant usage and grammatical weight, use of function words is nearly impossible to manipulate and thus will uncover motives, personality and psychological processes more accurately than analysis of the content (Pennebaker 2013). Using computational tools in analyzing function words bears further advantages. Firstly, people's poor awareness of function words is not restricted to their own language. The listener doesn't focus on function word composition, and therefore is unable to rate usage. Hence, computational pattern matching can reveal findings not attainable by human judges. Secondly, less than 0.04% of an average persons' vocabulary are function words (C. Chung and Pennebaker 2007). At the same time, they make up more than half of daily language. Consequently, function-word based analyses are well-situated to reveal latent individual states. All in all, the function word's importance on psychological findings justifies the application of the simpler dictionary-based approaches wherever emphasis is set on personal traits.

[18] Arabic, Chinese, Dutch, English, French, German, Italian, Portuguese, Russian, Serbian, Spanish, Turkish

Given its flexibility, ease of use, and localization, LIWC has been applied as the sentiment analysis toolkit of choice in many social indicators (e.g. happiness, characterizing network relationships, and opinion mining) studies (Lin and Qiu 2013; Niederhoffer and Pennebaker 2002; Ott et al. 2011; Pennebaker, Mehl, and Niederhoffer 2003). As such, numerous social benchmarks have been established and validated in cross-cultural and linguistic arenas. A summary of the most robust findings are listed below.

Happiness and Well-being

LIWC studies have demonstrated its capability in capturing two different dimensions of the happiness construct understood in the terms of psychological well-being (see Chapter 2.1 for an overview). In terms of the construct positive emotion, the study by (J. T. Hancock, Landrigan, and Silver 2007) researched which language dimensions shift based on whether the writer experiences positive emotion and is in a happy mood, or is situated in a context evoking negative emotion (Hancock et al., 2007). Intuitively, positive affection was found to score higher in the positive situation, and negative affection for the negative situation, respectively. This study isolated the LIWC categories ‘positive feeling’ and ‘negative feeling.’ Furthermore, participants in negative emotion employed negations more frequently while communicating. LIWC results of positive and negative emotion words were found to correspond with human ratings of text samples, thus proving its suitability for automated valence detection of positive emotion (Alpers et al. 2005, 370)

In accordance with its psychological origin, there has been much research on mental health assessment with LIWC dimensions. Whilst not as central to general community analysis, positive functioning and characteristics are important factors of well-being (Huppert and So 2013). Rude and colleagues revealed that people draw their attentional focus to themselves, when being in physical or emotional critical situations (Rude, Gortner, and Pennebaker 2004). They also use slightly more negatively valence words. Surprisingly an increase in first-person singular use was found to be a better marker for depression than emotion categories from the dictionary. Similarly, the usage of categories associated with higher cognitive complexity was significantly related to positive psychological functioning (Pennebaker, Mayne, and Francsis 1997). LIWC tracks these structures with numerable dimensions: ‘cognitive mechanism’, ‘cause’, ‘exclusion’, ‘negate’ and ‘prepositions’ are some examples showing increasing scores when complex processes accumulate.

Communal Belongingness and Social Communication

The existence of positive relationships and feeling of belongingness represents a further significant influence on well-being. (Baumeister and Leary 1995) describe the wish to belong as a basic human need, impacting well-being and health if not fulfilled. Belongingness describes the existence of interpersonal bonds providing the feeling of affective concern and stability. The need for belongingness is so critical, that total absence can be detected as a common driver in suicide attempts (Joiner Jr. 2005). Communal belongingness, as used in this work, refers to the ability of community members to feel being a valued part of and identifying themselves with the community (well in line with the micro-level of assessment of TSR). Communal belongingness is a valuable social indicator by which to describe communities.

Among the several LIWC categories pointing to belongingness, frequency of first person plural pronouns is a powerful indicator. An investigation found that internet chat room data four weeks after Diana, former Princess of Wales tragically died in a car accident, registered sudden and significant increases of the category ‘we’ (Stone and Pennebaker 2002). This finding coincides with Joiner et al., stating that in times of national tragedies suicide rates drop due to an increasing sense of belongingness within the community (Joiner, Hollar, and Van Orden 2006, 182). Another suicide study comparing text samples of suicide attempters and completers detected that the LIWC category ‘inclusion’ (e.g. with, include) is an effective way to measure belongingness. This is especially effective when contrasting inclusive words with the category ‘exclusion.’ Finally, LIWC offers a supra-category named ‘social processes,’ comprised of a diverse set of word groups to characterize communal belongingness.

Social communication also allows for determining status in terms of writers’ social hierarchy. Whilst high-status individuals refer frequently to other people (e.g. category ‘other’) low-status members tend to be self-focused and use tentative language (Tausczik and Pennebaker 2010). The authors also described the feature of linguistic immersion concerning emotion. They based this term on the results of a study dealing with women in abusive relationships (Holmes et al. 2007). There it was found that women used statistically significant more positive and negative emotion words when experienced pain was higher. It is intuitive to assume that, in general, adding emotion to communication depicts a deeper commitment to the subject, whereas formal and superficial descriptions lack emotive words.

Linguistic Accommodation

One basic requirement for LIWC being a usable tool is its ability to detect individual differences in language use. This potential was affirmed with the first study results (Pennebaker and King 1999). Yet, in mutual communication people frequently tend to converge their linguistic styles to promote social approval and communication efficiency (Niederhoffer and Pennebaker 2002, 339). This process is referred to as ‘Linguistic Accommodation’, ‘Linguistic Style Matching’ or ‘Linguistic Mimicry’ and is closely linked to the Communication Accommodation Theory (Giles, Coupland, and Coupland 1991). Several LIWC studies have elaborately researched this phenomenon and resulted that even in online chat rooms where stranger interact, mutual language adaption could be detected after several minutes and writing turns (Niederhoffer and Pennebaker 2002; Gonzales, Hancock, and Pennebaker 2010). Accommodation influenced word counts, emotive words, prevailing tense, complexity and many more. It was further revealed that intensity of adapting is not influenced by mutual liking, but rather by the degree of engagement to the conversation. That means a superficially friendly discussion will be more likely to depict individual differences than serious disputes. LIWC is considered well-suited for accommodation analysis, as linguistic mimicry represents a subconscious process, just as the function words LIWC focuses on (Gonzales, Hancock, and Pennebaker 2010). Obviously, subconscious partnership interest strongly increases degree of engagement, again supporting linguistic style matching.

Deception

People are considered to be the gold standard assessors of emotion and sentiment, and even people often have difficulties in detecting written deception (Ott et al. 2011). As a result automated lie detection is a fascinating research area as it goes beyond people's natural capability, and has innumerable practical and research use cases. One mechanism that has been applied to detect false stories is occurrence of logical mistakes and inconsistencies, i.e. high complexity and topic information is required (J. Hancock 2007). Researchers have hypothesized that people who are actively engaged in deception additionally differ in the way they formulate the text. Whereas the lie constructor has some potential to control the story to pretend sincerity, subconscious language patterns (e.g. function words) may be affected when actively establishing an event instead of reciting it from memory (Newman et al. 2003). Newman executed a deception study with LIWC, instructing participants to write each an English text excerpt in support of and denial of abortion, presenting both views as if they were the own opinion. Across studies with different media input (elicited written statements, elicited typed statements, video-transcribed statements, email, micro-bogs) it was revealed that liars:

1) Used less first person singular pronouns,

2) Expressed more 'negative emotion',

3) Used less complex terms.

The deceptive text samples reflected the missing personal relation to the story by their decreased use of first person singular references ('I'). Previous literature on deception further detected the intention of liars to dissociate themselves from the lie, experiencing a bad conscience (Newman et al. 2003). Tension and guilt are the explanatory variables for the higher usage of negative emotion. Furthermore, the required cognitive resources to deceive somebody reduces comfort in adding structural complexity and results in a shift to simple, descriptive verbs. Hence the score for 'exclusion' dropped among liars and the category 'motion', consisting of simple verbs, showed an increasing frequency.

With help of these findings LIWC was able to correctly uncover deceptive text samples with 67% accuracy. In contrast, human judges only classified 52% of the same data correctly, basically the performance of guessing (Newman et al. 2003). Obviously the critical difference in language between false and true stories does not only leak through the tellers' subconscious without awareness, but is also hard to be captured by human judges, as they focus more on the content of stories than observing these hidden subtleties

5.3 Applications of TSR

TSR aims at measuring and improving well-being in connection with provision of services. Movement towards this goal requires, among many other things,

> 1) Identifying and understanding the variables that affect well-being in conjunction to the service experience, and
>
> 2) Obtaining said data.

This chapter addresses that research gap by proposing an extended framework based on (Anderson et al. 2013) for the configuration and measurement of these variables, along with a comparison of existing data sources at the national and international levels, and possible methodologies to the collection of personalized data, namely gamification and text analytics (**RQ 1.2**).

The foundational argument to this thesis is that currently missing are the tools and indicators needed for designing TSR for individuals in the service pyramid. An obvious and important use of currently existing data sources is to have them serve as benchmarks for the coming analyses. There are two such modes of use. The first is for validating new instruments to be developed by TSR scholars, as addressed in Section 3.2. Existing questionnaires and other instruments (see Chapter 2 for a review of well-being measurement instruments), as well as the data collected with them can be used in designing new instruments and in testing them, e.g., for application in serious games (**RQ 2.1**). A second valuable role of these data is to serve as comparison points for studies done at smaller institutions or regions, e.g., constituents of a given community (**RQ 2.4**). Very often, targets of TSR will be particular institutions (government agencies, commercial firms, NGOs, etc.) that are on a much smaller scale than the most widely-used, macro level surveys. Data targeted at a particular institution will be able to compare the effect of the institution against that of the larger society, or in the formalization of value co-creation between providers and consumers.

In summary, this chapter addresses both of the listed requirements by surveying existing literature and exemplary application contexts (gamification and text analytics), and existing data collection efforts and archives that are relevant to TSR and that have high-quality data publicly available (e.g, the GSS and Better Life Initiative). A third contribution comes from the delineation of well-being terminology and applications in a way which moves towards a taxonomy of well-being measurement (**RQ 2.1**). Together, these sources of findings constitute something of a map of (some) resources—both of a conceptual nature as well as hard data—available to the TSR community. Building on a wealth of existing knowledge and attending to new developments, TSR is poised to contribute enormously to fostering well-being.

On Defining Well-being for Progressive Community Management

Chapters 2 and 3 address the first two research questions of this thesis. **RQ 1.1** is addressed in detail theoretically, defining the attributes necessary for the use of well-being as an indicator. To do this, well-being was delineated and defined from three viewpoints: economics, philosophy, and psychology. A working definition of well-being for this thesis, introduced as Human Flourishing, was provided. Then the attributes of transformative service research are introduced as macro-, meso-, and micro- service interactions. Macro-interactions refer to the environment in which an individual exists; meso-interactions represent the self-perception of the individual's place in that environment. Micro-interactions, by far the least addressed and most difficult area to measure, are the foundational psychological underpinnings which shade the view of the individual in a given situation. Each aspect is necessary to consider in TSR.

In describing the necessary considerations of these three service interactions, **RQ 1.2** is partially addressed. Data collection for well-being has until now been largely offline with representative populations via surveys and interviews. Unaddressed is the replication of such studies in online fora. Also unaddressed is the granularity of well-being studies, which is to say, what occurs when well-being is applied as an indicator for non-national scale assessment? In measuring the micro-interaction of TSR, online social media promises to provide abundant and varied data types from which to analyze personal well-being. The mechanisms and supporting technologies of serious gaming and text analytics and their respective methodologies are discussed as two particularly promising aspects of the digitalization of daily life from which to measure well-being (**RQ 2.1**). Gamification allows the elicitation of well-being in a stated preference scenario; text and sentiment analysis allow the reconstruction of revealed preference via actual behaviors and expressions. As such, this lays the groundwork for applied assessments of gamification and text and sentiment analyses based on online social media in the assessment of well-being for use in transformative service research.

Part III.
Applied Well-being Measurement in Institutions

Chapter IV BeWell: A Game of You on Facebook

"How to gain, how to keep, how to recover happiness is in fact for most men at all times the secret motive for all they do."

William James, Varieties of Religion Experience (1902)

Responsibly collected well-being data can drive proactive institutional management. Integrating the well-being data of individuals, and their history into a TSR application has practical implications that are directly applicable to institutional management: They can help managing complex communities or institutions beyond the less precise instruments employed today. The relationship between personal and communal well-being is the fundamental base for TSR. At the basest level, communities are made by personal (meso-level) interactions with other individuals, groups, institutions and events. (Micro-level) perceptions of these interactions drive personal perceptions of well-being, which among other indicators is a (macro-level) predictor of social cohesion (Thinley 2011), a necessary condition for progressive communities. Notably, it can be assumed that a significant drop in the projected long-term expectation of an individual's or a community's well-being is a clear indicator that calls a community manager to action – and provides a strategic advantage to those community managers that are in possession of a tool, in the best case online, that enables the evaluation of such measures (Davies 1962). The effectiveness of TSR depends on suitable data: It must reliably reveal the actual well-being level of individuals as a comparable measure and it must represent such levels timely distinct, yet granular enough to enable the construction of trends and their analysis. Together, this would allow for the precise tracking of well-being over time. For the purposes of this work, "institution" and "community" are used synonymously.

Today's institutional indicators, notably turnover rates, performance assessments, and absentee tracking are no longer adequate, as they do not possess the multidimensional aspects and conditional factors needed to manage institutions. The challenge facing the management of on- and offline communities, as well as the overall success and health of institutions, is to identify fitting well-being indicators utilized in an appropriate method (Ahn et al. 2011; Anderson et al. 2013). Constituents, decision makers, stakeholders as well as human resource divisions lack adequate measures to determine the state of psychological or social health in their institution (Harter, Schmidt, and Keyes 2003; NEF 2009; Grant, Christianson, and Price 2007). This knowledge gap hinders decision and policy makers in implementing TSR. To circumvent potentially significant gaps in knowledge, digital well-being measurement is needed as a “best practice” mechanism for tracking thriving on- and offline communities. The challenges in accordance with Research Question 1.2 are twofold:

1) A mechanism for well-being assessment has to be designed, and
2) A transparent yet secure data collector needs to be developed and tested.

This work explores the possibilities of the use of gamification on social network platforms for individually elicited, real time well-being data in order to populate a TSR application. Firstly, a progressively larger series of surveys are implemented online as pilots; secondly, several machine learning algorithms are applied to data collected via surveys in order to provide insights regarding the dependencies between personal well-being (dependent variable) and personality as well as demographics (independent variables). Thirdly, gamification and its mechanisms are evaluated to address issues revolving around participation incentives using techniques in social network propagation. This gamification lead to the development and prototyping of **BeWell: A Game of You on Facebook**, a Facebook-based app for well-being measurement.

This chapter is an exploration and extension of the collective works (Hall, Caton, and Weinhardt., 2013; Hall, Glanz, Caton, and Weinhardt, 2013; Hall et al., 2012) as well as the working paper (Wilckens and Hall 2015). It starts with a description of a pilot study, (Section 4.1) which reviews the validity of well-being survey items collected via online social media. Section 4.2 reviews two feasibility studies of the use of well-being for progressive community management and evaluates the statistical methods and machine learning algorithms used as the prediction engine of the eventual game. The prototype **BeWell**: **A Game of You on Facebook** is introduced (Section 4.3), then evaluated along with directions for future work in the measurement and assessment of personal well-being and online participation (Section 4.4).

4.1 Application of Design Science to BeWell

Context-dependency of the effectiveness of gamification methods is repeatedly expressed in scientific literature (see Section 3.2.3 for an overview of the literature). In this section, those incentive factors are introduced and discussed as they pertain to the iterations of this research. Further, four dimensions that served to analyze the incentive factors regarding their dependencies among each other and prerequisites will be presented. The original aims and requirements were to identify a subset of incentive factors whose effectiveness could be verified under laboratory conditions; it became apparent that a more sophisticated approach would be needed. To this extent, Design Science was employed.

4.1.1 On the Suitability of Design Science as a Method

From a methodological perspective, the systematic process of design that is Design Science (Hevner et al. 2004; Peffers et al. 2007; Winter 2008) is well suited to address the research questions introduced at the beginning of this chapter. The usage of gamification elements to create an application to measure Human Flourishing is a novel approach to be addressed with an instantiation of an artifact. The application of two previously unused mechanisms together and the various interactions and context dependencies thereof need to be investigated in a manner that allows for rigorous evaluation. Prior knowledge on the interaction between Human Flourishing measurements and gamification is not available. Scientific literature emphasizes the context dependency for the efficient application of gamification elements in many respects (Antin and Churchill 2011; Deterding 2011; Siegel 2012; Vassileva 2012) therefore it is difficult to deduct findings from other gamified applications (see Seciton 3.2.3 for examples). The same is true for the purposeful, context-dependent inclusion of basic gamification elements from the knowledge base. Here, Design Science with its explicit expectation of creative contribution fits well. Finally, the Design Cycle advocated by this methodology is well suited to the research conducted by this thesis, as seen in Figure 4.1.

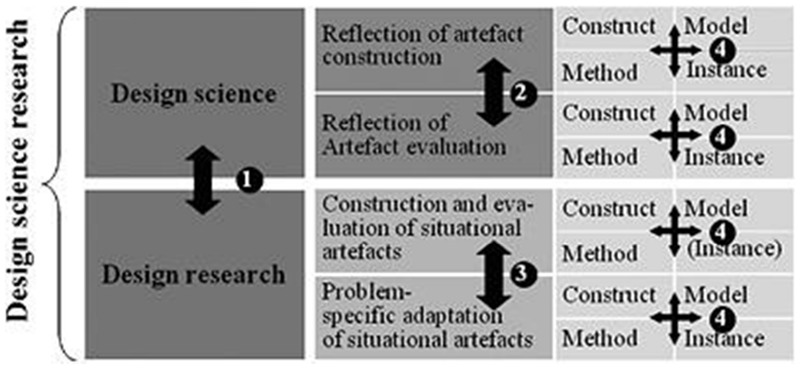


Figure 4.1: Design Science research cycle of (Winter 2008)

The ‘Construct-Model-Method-Instance’ cycle provides for the means necessary to iteratively create and improve such an artifact. This way, the lack of prior specification can be handled through constantly bringing in findings from related literature, combined with creative input from the researcher and continued evaluation, e.g. by test users.

Design Science enables the iterative reflection and construction of an artifact to define, develop, demonstrate, and evaluate in a way that is tailored to the exploratory nature of the research at hand that is scientifically sound (Peffers et al. 2007). In case of this thesis, it means creating an artifact that investigates on the identified, relevant problem of measuring well-being as a serious game. Therefore, it must reliably collect truthful well-being data and incentivize its users to continuously provide this data. The artifact needs to be developed in a way that inherently allows for its evaluation, this way being suitable to provide a solid answer to the research questions. It needs to be demonstrated that the artifact fulfills its purpose by setting it into a fitting application context. Design Science is distinct from general system building not only because it sets it emphasis on the creation of innovative artifacts, but it also inherently considers the evaluation of results.

In accordance with the principles of Design Science, this thesis introduced four iterations of online and progressively gamified surveys. The iterations also had progressively more observations per participant.

1) An initial pilot study testing cross-sectional Human Flourishing reporting online (n=174), released on Facebook,
2) A longitudinal survey with four observation points evaluating Human Flourishing and personality in 2013 (n=85), announced on Facebook and email,
3) A larger scale instance of the second iteration (n=343) in 2014, announced on Facebook and email,
4) A fully gamified proof of concept (**BeWell POC**) iteration evaluating Human Flourishing and personality (n= 121), released on Facebook.

All iterations were introduced and completed between July 2012 and March 2014. Iterations two and three were held consistently over the four Wednesdays that occur in February over two years to allow for consistency in reporting. Wednesdays were chosen to avoid spikes and dips in happiness due to the occurrence or ending of weekends. A test question “Take a look out of the window. How is the weather today?” was implemented at the start of each survey with a free-text box. This was used to both filter unserious respondents, and the mitigate the effect of the weather on mood (for a discussion of how to mitigate the impact of weather on subjective states, see (N. Schwartz and Clore 1983; Kahneman and Krueger 2006, 6)). The coming section discusses the design issues central to the application of gamification to well-being measurement.

4.1.2 Identification of Incentive Factors

Possible incentive factors that could be applied to the envisioned, final version of **BeWell: A Game of You on Facebook** were identified and clustered into different groups. The groups identified are “Inherent, nearly-exclusive incentives of BeWell” consisting of incentive factors that deal with the calculation, charting, and different forms of comparison of well-being data; “Further intrinsically motivated incentives” consisting of items that link to the helpfulness or demand for self-expression of the user; “Basic game mechanics” that describe an supportive application environment and point system; and “Social mechanics” that contain incentives designed to take advantage of the motivational effects of direct user-to-user interaction.

Having identified a rather high number of factors in the literature (Section 3.2.3), each was examined regarding the four dimensions Implementability, Context, Testability, and Miscellaneous described in the below Table 4.1.

Table 4.1: Dimension of incentivization in serious games

Dimension	**Description**
Implementability	Are there serious constraints that could hinder the implementation of the proposed incentive factor?
Context	Is the functioning of an incentive likely to depend on a “real situation” that could not be simulated in a laboratory-like setting?
Testability	Does it seem demanding to test an incentive factor because it would require a high amount of data or time, including the need of multiple sessions on different days (with the same test user(s))?
Miscellaneous	Other possibly problematic points of interest, e.g., Does it seem likely that the usage of an incentive factor could interfere with the usage of another incentive factor? Does it seem likely that the usage of an incentive factor could interfere with the collection of unaffected (truthful) well-being data or with the basic protection of the users’ privacy?

In short, the context specificity of most incentive factors is indeed present, as well as interdependencies amongst the factors. Given the above considerations the creation of a proof of concept implementation that implements a plausible subset of the identified possible incentives was devised. That subset was chosen in a way to provide the necessary overall, interconnected context of well-being gamification. Additionally, testing should be done under realistic conditions, i.e. the proof of concept implementation should be released to Facebook.

4.1.3 Objectives of the Solution

The proof of concept implementation BeWell POC has a variety of objectives. That is caused by the fact that it bridges several areas of knowledge, namely bringing together gamification with well-being measuring as a web application while providing for built-in evaluation. The objectives are framed through an iterative process with multiple repetitions and refinements. It contains application of findings from literature review, the purposeful inclusion of success measures, building early proof of concept implementations, review by testers, comparison with other gamified applications, and the adaptation of best-practices. This way the objectives evolved from a rather small, mockup-based first vision to a more sophisticated, rather feature-rich vision of BeWell POC.

BeWell POC supports experimental setups and the collection and storage of an extended set of data. The data collected generally allows for being represented and analyzed in a variety of ways, including statistical methods (discussed in Section 4.3). BeWell POC focuses on the effectiveness of certain gamification incentives and the meaningfulness of the flourishing-related data provided by its users. In the sense of Design Science, it is planned to be a step within the overall iterative process to construct a gamified application for measuring well-being.

Primary deduction from gamification

Besides motivating the decision to move on with the development of a proof of concept implementation following Design Science, the identification and examination of possible incentive factors also produced the following starting point for defining its objectives: BeWell POC needs to mimic a realistic environment for gamified well-being, including a basic "gaming platform" with the implementation of an adequately high number of additional, interconnected incentives. Additionally, BeWell POC should not use incentives that allow for the comparison of Human Flourishing scores as this form of comparison could have harmful influences on truthful reporting by some users (Ryff and Keyes 1995; Guven and Sørensen 2012). A constantly visible Human Flourishing of a specific social sub-network or of specific users could be (mis)interpreted to be a "reference score." This could cause several reactions. It could be possible that a user with a non-average score experiences the (subconscious) urge to manipulate his reporting behavior to get closer to the reference score (Utz, Tanis, and Vermeulen 2012). While one can imagine that this is especially true for sub-scorers, depending on a user's personality and/or social context also an adoption in the other direction could occur for high-scorers. Further, one could think of a behavior that aims at keeping a certain distance to the average or specific "benchmark" score (Dixon 2011). Just as well, users (overly) convinced of themselves could (subconsciously) regard it as necessary for their self-image to have scores over average or to "perform better" than specific users selected to benchmark against (Guven and Sørensen 2012). Further research can be undertaken regarding these suspicions, but the current iteration of BeWell POC will concentrate on the basic applicability of gamification to well-being measuring.

The initial selection of this additional incentives was basically inspired by the list of possible incentives for "BeWell: A game of you on Facebook" (see left column of Table 8 in the first sub-chapter of the Appendix). Over the course of developing and extending BeWell POC to its release version, most of those incentives were implemented in some form. This is particularly the case for the groups "basic game mechanics" (14 - 25) and "social mechanics" (26 - 29). The incentives related to knowing one's own well-being level and its evolution/history were the only ones from the group "inherent, nearly-exclusive benefits / incentives of BeWell" that were implemented by BeWell POC. This is due to the fact that incentives that allow for the comparison of Human Flourishing Scores were deliberately excluded. In an attempt to represent the group "further intrinsically motivated incentives", badges were designed in the two distinct and leveled flavors "Scientific Advance" and "Better World".

Iterative refinement and final scope

Over the course of development, the primary objectives were refined and extended in the sense of Design Science. For building a foundation, a functional, error tolerant Facebook application is implemented and equipped with a configurable “Question Engine” that allows for reliable and varied data collection with support of different question types (slider-, pictogram-, and text-based). An algorithm to calculate the Human Flourishing Score of the user is designed and implemented based on the calculation of Human Flourishing found in Equation 2.3. Finally, a subset of additional gamification incentives is provided within the application to create a realistic environment for gamified well-being.

From the necessity to measure the success of BeWell POC and due to the implications that the creation of a potentially far-spreading Facebook application has, additional or supportive objectives were deducted. Tracking capabilities that allow recognizing application errors, how the application is used by participants, and the use of incentive mechanisms by participants were implemented. Also, a built-in questionnaire function was implemented. All user-provided and tracking-related data was stored in a way that is privacy sensitive and allows for versatile analytics. Basic demographic data about the user, including gender, age, country of residence, and highest successfully completed level of education was collected in a way that allows for its change by the user to accommodate non-truthful reporting on Facebook’s About Me section.

Further, recognizing the personal nature of the data collected basic protection of the user’s privacy is to be supported. Being a web application, counter measures against a basic set of well-known attacking methods in the web environment must be included. Finally, BeWell POC was localized in English and German, being the most prevalent languages within the expected user base. This was meant to lower entrance barriers and to reduce the risk of false reporting because of language-dependent misunderstandings.

4.2 Well-being in Community Management

To test well-being’s reliability when collected via online social media and the general willingness of participants to participate in TSR-like data collection exercises, a pilot study was conducted in July 2012. Using the definition of (Huppert and So 2013) the pilot looked at the ten basic items of Human Flourishing (see Section 2.1.4). The presence of positive emotion, competence (f_1), meaning (f_2), engagement (f_3), positive relationships (f_4), emotional stability (c_1), self-esteem (c_2), optimism (c_3), resilience (c_4), and vitality (c_5), and demographic questions were asked in an online survey format (See Appendix I for survey details). The survey applied (Huppert and So 2013)’s Human Flourishing survey, as addressed and calculated in Section 2.1.4, Equation 2.3.

The responses showed high validity and a reasonable sampling of typical online social media consumer (For demographic information, review the work (Hampton et al. 2011)). This is a positive reflection on the ability of serious games to elicit data for the purposes of TSR. 174 respondents completed the survey. Of these, 22.4% answered in German and 77.6% answered in English. Respondents' self-reported locations in North America (78), Europe (75), Asia (12) and Africa (1), with eight declinations to respond. 94 respondents self-reported their gender as 'Female', 74 as 'Male' and six respondents declined to report a gender. This gave a slightly higher response percentage from women (54%) than men (42.5%), a potential selection bias issue. Self-reported educational attainment shows 130 of the respondents hold at least a Bachelor's degree. The age distribution shows that most respondents are between 20 and 40 years old (Figure 4.2(a)).

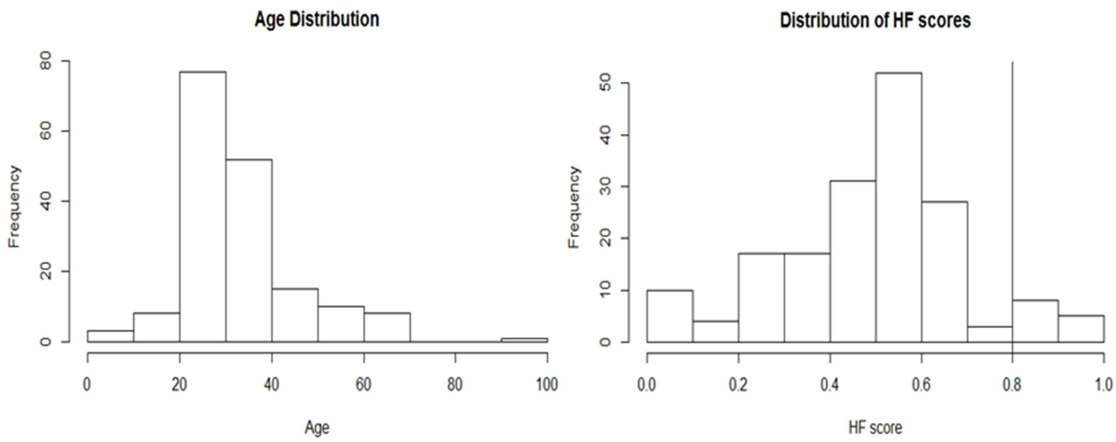


Figure 4.2: (a) Age distribution of the survey respondents, (b) Histogram of Human Flourishing scores

Based on the formula of Human Flourishing (Equation 2.3), a raw, human flourishing score (HFS) was calculated. The distribution of the HFS's is shown in Figure 4.2(b) as a histogram, where the vertical line shows the cutoff value of 80% of the maximum achievable score, which was used by Huppert and So to distinguish between highly flourishing and the rest of the population in their initial study. Calculated at the .80 threshold, 13 participants (7%) would fit Huppert and So's definition of being highly flourishing. This is considerably higher than the 7.3% reported in (Huppert and So 2013, 848), likely due to the differences in geographic regions sampled in the two populations (discussed further below). The mean value of HF is 0.49, with a standard deviation (SD) of 0.20.

Non-parametric Mann-Whitney-U tests (Wilcoxon rank-sum tests) and Kruskal-Wallis tests revealed that there was no statistically significant difference between HF based on gender, age or education. However, a Wilcoxon test on the difference between HF reported from North America and Europe, (as well as a Kruskal-Wallis test between North America, Europe, and Asia) revealed statistically significant differences at the 1% level. That North

Americans self-report higher well-being levels than Europeans is well-established (Okulicz-kozaryn 2011); it should be noted that self-reporting well-being and actual experience of well-being are not to be conflated. It would be incorrect to say that North Americans are happier than Europeans.

Table 4.2: Spearman's rho of Human Flourishing with significance levels (***: p<0.001, **: p<0.01, *: p<0.05)

	Positive Emotion	Competence	Engagement	Positive Relationships	Meaning	Emotional Stability	Optimism	Resilience	Self Esteem	Vitality
Positive Emotion	1.00									
Competence	0.60 ***	1.00								
Engagement	0.28 ***	0.27 ***	1.00							
Positive Relationships	0.36 ***	0.30 ***	0.22 **	1.00						
Meaning	0.60 ***	0.66 ***	0.31 ***	0.33 ***	1.00					
Emotional Stability	0.49 ***	0.35 ***	0.17 *	0.16 *	0.32 ***	1.00				
Optimism	0.46 ***	0.43 ***	0.32 ***	0.36 ***	0.50 ***	0.34 ***	1.00			
Resilience	0.19 *	0.13	0.09	0.08	0.08	0.18 *	0.30 ***	1.00		
Self Esteem	0.51 ***	0.37 ***	0.39 ***	0.30 ***	0.46 ***	0.44 ***	0.57 ***	0.31 ***	1.00	
Vitality	0.49 ***	0.37 ***	0.35 ***	0.24 **	0.45 ***	0.53 ***	0.32 ***	0.19 *	0.53 ***	1.00

Considering the correlation values of the ten items of Human Flourishing (see the Spearman correlation values in Table 4.2) in the pilot study, there is a positive correlation between all items with the exception of resilience. This is not surprising based on the way that the HF is calculated. It is found that none of the input variables display multicollinearity, the status of having two or more items that are highly correlated (meaning that items, combined or not, could linearly predict the others) (Belsley 1991). However, these correlations do not replicate the Spearmen's correlations found in the initial study (Huppert and So 2013), likely due to the difference in sample size. The pilot study showed that well-being can be reliably recorded online, and that public propagation would be a feasible mechanism to gather TSR data in the future. The initial use case verifies the suitability of this data to be used in in support of TSR.

The pilot was however based on a cross-sectional study. To be further investigated is the scalability of such a system in a longitudinal as opposed to cross-sectional study. Such a measured approach is in line with the iterative requirements of Design Science.

Second and Third Iterations in the Design Cycle

Human Flourishing values are subsequently investigated as a prediction problem- that is, can well-being be predicted (individually or in subgroups), when psychometrics and demographics are considered in a longitudinal scenario? To approach this, the second and third iteration of the online survey with four sequential questionnaires and an overall number of 126 questions was launched (Figure 4.3 reveals the variable structure; see Appendix I for the full listing of items). The second iteration was completed in February 2013 and the third in February 2014. These psychometric tests have low variance over time, and thus can be tested once and still are considered valid for the length of this one-month survey (Huppert & So, 2013; John, Donahue, & Kentle, 1991; Schwartz et al., 2002). Respondents were given the option to review their results at the end of the four weeks. The 2013 iteration generated a dataset of 85 participants during a four weeks period in February 2013. The February 2014 iteration expanded to 343 participants.

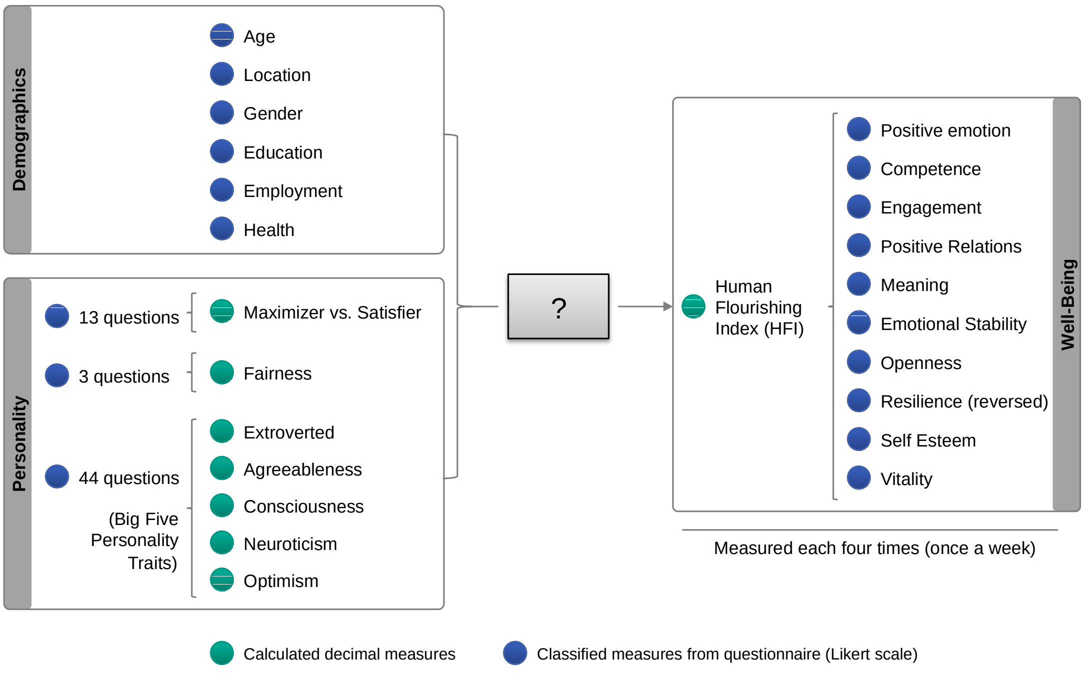


Figure 4.3: Independent and dependent variables in a well-being prediction scenario (represented as a question mark)

The participants were asked by email to answer one questionnaire each Wednesday in the month of February, 2013. Of 85 initial respondents from the first questionnaire in week one 66 participants completed all four questionnaires entirely. Nine participants aborted after week two and another four participants after week three. From seven participants only single values are missing, with an overall loss of 14% of the participants across four weeks. Self-reported gender revealed a 50-50% female-male split, with one non-response. Three participants who completed the surveys self-reported being located in Asia; 22 from the United States; and 34 self-reported locations within Europe, with four declining to respond. 78% self-report being age 35 or under. 85% of respondents reported being currently employed. 81% of the respondents self-reported completing at least a master's degree. 86% of respondents refer to themselves as "moderately healthy" or "very healthy."

Due to the small sample size, it was decided to repeat the survey during February 2014, exactly one year after the first series in order to avoid seasonal influences. An additional dataset with 343 respondents for the first questionnaire was generated. The questions and the setting for the four questionnaires were identical to the one in 2013. 296 participants completed all four questionnaires. While still small, this sample is meritorious of application of advanced statistical techniques. In total 13 independent variables and 4 Human Flourishing score (HFS) data points were calculated per participant and standardized with minimum zero and maximum one for the descriptive analyses. In order to perform machine learning algorithms the data is further normalized to zero mean and SD of one per variable. These include six demographics and seven psychometric measures, calculated upon single items. If one of the 13 input dimensions was missing, or a subject reported less than three HFS data points were available, the subject's information was eliminated from the dataset.

4.2.1 On Survey Item Suitability

A principal components analysis (PCA) was completed with the February 2013 iteration, considering the survey items proposed and validated by: (Huppert and So 2013; John, Donahue, and Kentle 1991; B. Schwartz et al. 2002; Schmitt and Do 1999). Inspection of the correlation matrix showed all variables had at least one correlation coefficient greater than 0.3, meaning PCA is a valid data reduction method (Kaiser 1970). The overall Kaiser-Meyer-Oklin (KMO) measure was 0.818 with most individual KMO measures all greater than 0.7, classifications of 'middling' to 'meritorious' according to (Kaiser 1970). Exceptions here are 'Optimism' at 0.621; Maximizing at 0.499; Fairness at 0.352; and Engagement at 0.667. In accordance with the recommendations of Kaiser, these items are retained but closely observed. Bartlett's Test of Sphericity was statistically significant ($p < .0005$) indicating that the data was likely factorizable (Gleser 1966).

PCA revealed five components that had eigenvalues greater than one and which explained 37.4%, 9.1%, 8.3%, 7.4%, and 6.0% of the total variance, respectively. Visual inspection of the scree plot indicated that all five components should be retained (Chou and Wang 2010). In addition, a five-component solution met the interpretability criterion. As such, five components were retained. It must be noted here that in line with the KMO results, the fourth and fifth factor are weakly clustered with other items.

Table 4.3: Component transformation matrix

	1	2	3	4	5
1	.815	.489	.304	.061	-.023
2	-.037	-.437	.802	.136	.383
3	-.385	.560	.277	-.633	.245
4	.419	-.396	-.335	-.563	.488
5	-.103	.316	-.275	.510	.744

Extraction Method: Principle Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

The five-component solution explained 68.2% of the total variance. A Varimax orthogonal rotation with Kaiser normalization was employed to aid interpretability. The interpretation of the data was consistent with the personality attributes the questionnaire was designed to measure with strong loadings of well-being items on Component 1, personality items on Component 2, optimism items on Component 3, maximization items on Component 4, and fairness items on Component 5. Component loadings and communalities of the transformed solution are presented in Table 4.3.

4.2.2 *Data Descriptives*

Firstly, the similarity of the two datasets is assessed. The high percentage of explained variance indicates a larger deviation between participants than within each participants HFS trajectory (Table 4.4). This is an indication that individuals are by and large consistent in their reporting, though there are differences across individuals. This can also be found within the SDs (Table 4.5).

Table 4.4: Explained variance of weekly HFS by the HFS average

	HFS week 1	HFS week 2	HFS week 3	HFS week 4	Average
Weekly HFS variance explained by HFS average	79.96%	88.72%	86.21%	79.76%	83.66%

The averaged SD within each participant's HFS values (0.077) is 2.5 times smaller than the SD between participants averaged HFS value (0.1954). As shown in Table 4.4, the averaged HFS per participant accounts for 83.66% of the variance within the weekly HFS data, a significant increase from the pilot study.

Table 4.5: SD between and within participants' HFS trajectory

	2013 Dataset	2014 Dataset	Combined Data
Avg. $SD_{within\ particpant}$	0.0787	0.0765	0.0769
$SD_{between\ particpants}$	0.2035	0.1915	0.1954
Ratio	2.59	2.50	2.54

When considering the seven personality traits tested throughout the survey (sensitivity to fairness, maximization, extroversion, neuroticism, optimism, agreeableness, and conscientiousness), the results across subpopulations are much more varied than are found throughout the Human Flourishing items. This is encouraging, as the attributes here are a hypothetical basis of how the gamified survey predicts well-being based on subpopulations. An overview on the resulting data dimensionality is seen in Figure 4.2.

Figure 4.4 provides a descriptive impression of the HFS distribution in which data is sorted by the averaged HFS per participant, and reflects a reversed sigmoid distribution. The solid dark line indicates the averaged HFS per participant; the error bars cover each participant's single weekly values from minimum to maximum. The sample is well distributed over the whole well-being scale from zero to one with an average of 0.55 as presented in the density plot (Figure 4.5). The small peaks at zero and one result from special characteristics of the HFS, which has several input constellations leading to extremes at zero and one.

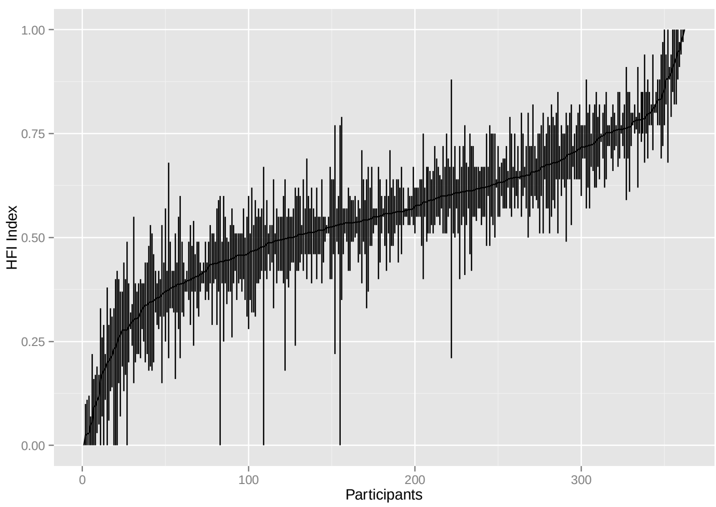


Figure 4.4: HFS distribution

For each individual HFS data point the hour of the day has been recorded, in order to control for possible influences caused by responses in the day or night. Except for a slight decrease in the late evening after midnight, no significant influence was observed. Moreover, the lower averages during nights are based on a few values with high variance only and are hence not further considered as standard.

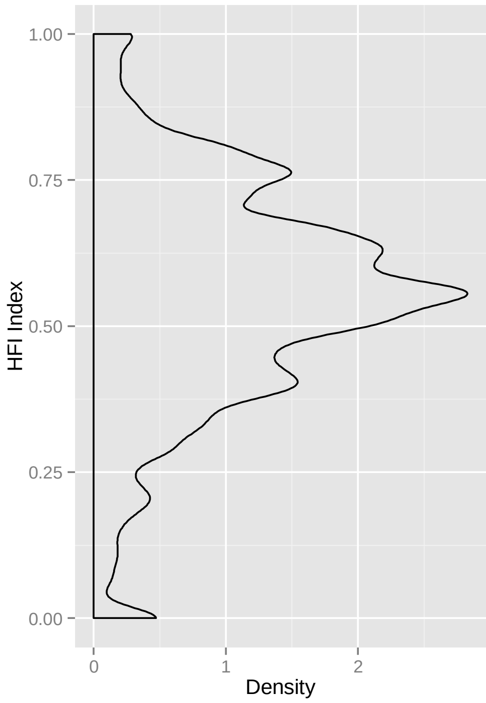


Figure 4.5: HFS density

In order to check for multicollinearity, a graphical representation of the correlation matrix for all variables in the dataset is given in Figure 4.6. It is found that none of the input variables are highly correlated to others. Additionally, the condition of the input matrix is 12.6, indicating weak dependencies (Belsley 1991). As a result, multicollinearity is not considered, indicating that multivariate models can be applied without previous feature reductions.

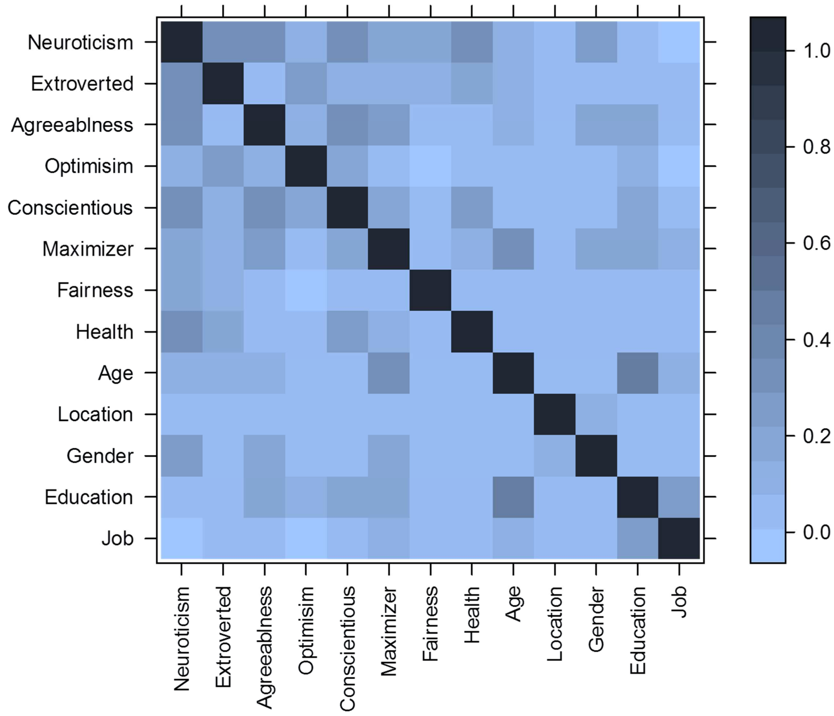


Figure 4.6: Correlation matrix (absolute values)

Overall, the items found in the employed surveys are found to be suitable to the task of assessing individuals' psychometrics. As the underlying structure of the data is factorizable without multicollinearity, it is also suitable for use in prediction problems.

4.3 Evaluation Methods of Well-being and Baseline Personality Traits

The data has several characteristics. It is sensitive, as it deals with personal standards and perceptions; it is noisy, due to the multi-layer collection method; and while correlation potential between the interplaying factors is possible, causation is nearly impossible to reach. The downside is however that there could be a very high amount of signal variance across and within people, making it a non-trivial classification problem. A high degree of computational analytics with a high degree of sensitivity is required to make well-being prediction feasible.

After calculating Human Flourishing, a multiple liner regression was modeled for predicting the Human Flourishing score as a dependent variable from the psychometric attributes. The assumptions of linearity, independence of errors, homoscedasticity, unusual points and normality of residuals were met (Nelder and Wedderburn 1972). The linear regression established certain psychometric traits could statistically significantly predict Human Flourishing, $F(13, 51) = 9.116$, $p < .0005$. Regression coefficients and standard errors can be found in Table 4.6.

Table 4.6: Results of a linear regression model, Human Flourishing and psychometric attributes

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model	β	Std. Error	Beta		
(Constant)	.194	.236		.822	.415
MS Scale	-.013	.021	-.061	-.640	.525
Extroversion	.070	.020	.346	3.568	.001***
Agreeableness	.057	.034	.187	1.703	.094
Conscientiousness	.003	.029	.031	.117	.907
Neuroticism	-.102	.023	-.452	-4.479	.000***
Openness	.015	.024	.060	.634	.529
Fairness	.041	.025	.149	1.613	.112

Dependent Variable: Mean Human Flourishing Score

With an R score of .727 and R Square of .528, the feasibility of making predictions of Human Flourishing is considered to be reasonably accurate. This is further confirmed by the results of an ANOVA on the linear model (Table 4.7) which confirms that at least one of the predictors has a highly significant correlation to Human Flourishing.

Table 4.7: Analysis of Variance (ANOVA), Human Flourishing and psychometric attributes

	Sum of Squares	df	Mean Square	F	Sig.
Regression	.899	7	.128	9.116	.000***
Residual	.803	57	.014		
Total	1.701	64			

Dependent Variable: Mean Human Flourishing Score; Predictors: (Constant), Fair Mean, MS Scale, Extroversion, Neuroticism, Openness, Conscientiousness, Agreeableness

Of the seven predictors, neuroticism and extroversion have the highest weight (discussed in detail below). Neuroticism is highly significant at the 0.001 level with a negative coefficient estimate. This indicates that higher levels of neuroticism predict lower flourishing levels. Extroversion is also highly significant at the 0.001 level with a positive coefficient estimate. This indicates high extroversion is predictive of high flourishing levels. The strength of these two relationships to overall Human Flourishing scores is notable, as it suggests that inferences about the population can be made.

4.3.1 Assessment of Predictive Models for Well-being Prediction

Important to the utilization of prediction well-being for community management is the assessment of the best performing model. To that extent, the generalized linear model (GLM) (a backbone of machine learning) and the machine learning algorithms from the kernel-smoothing,[19] neural network,[20] and feature selection[21] families were applied (Figure 4.7). Whilst interesting results were found across the different models, the best overall performance was found with the GLM, with close performance achieved with the local linear regression family. Linear Extreme Machine Learning meets the performance standards of GLM. However, GLM was selected as the benchmark due to its overall low complexity in comparison with linear Extreme Machine Learning. Overall performance considers both accuracy of prediction by observations and explained variance. This section explains the results of the GLM, and supplemental information of the performance metrics of can be found in Appendix II.

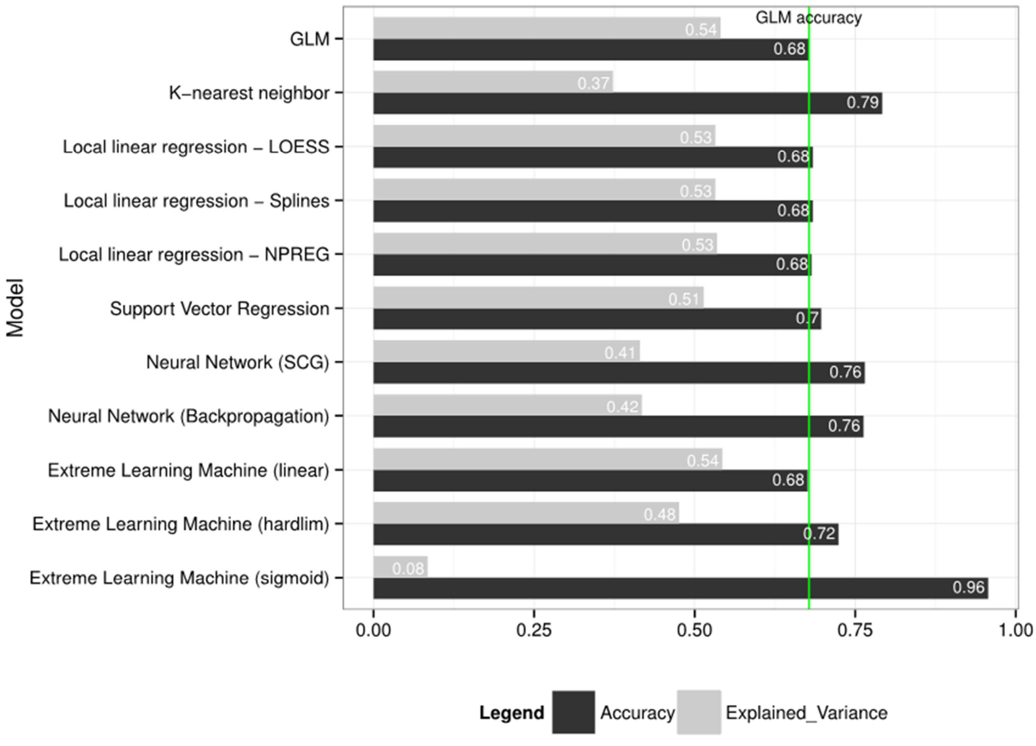


Figure 4.7: Accuracy comparison between deployed algorithms for well-being baseline prediction

[19] Including K-nearest neighbor, non-parametric regression, LOESS, Splines, and NPREG.
[20] Including Stuttgart Neural Network Simulator for R and Extreme Machine Learning.
[21] Including lasso and elastic net regression, and lazy lasso regression.

The GLM is an important benchmark for advanced machine-learning algorithms considering non-normal input variables. The GLM is a generalization of the standard linear regression that allows for non-normal distributed dependent variables (McCullagh 1984). Therefore, a GLM including all 13 predictors and the averaged HFS as dependent variable is conducted with 10 times repeated 10-fold cross-validation. Multi-fold cross-validation on has been proven to be a valid bias-reduction measure (Zhang 1993). The GLM results in an R^2 of 0.54 and a root-mean-square-error (RMSE)[22] of 0.68. The non-cross-validated standard linear model fitted to the entire dataset reaches an only slightly better RMSE of 0.66, so that over-fitting is an unfounded concern for this model. The results are equal for both combined datasets: for 2013 a RMSE = 0.67 and for 2014 a RMSE = 0.69 is achieved.

Compared to the SD of the averaged HFS (normalized to SD = 1) the GLM predicts the independent variable 32% better than a simple average prediction. Each predictor's importance, measured by the absolute value of the t-statistic, is given in Figure 4.8.

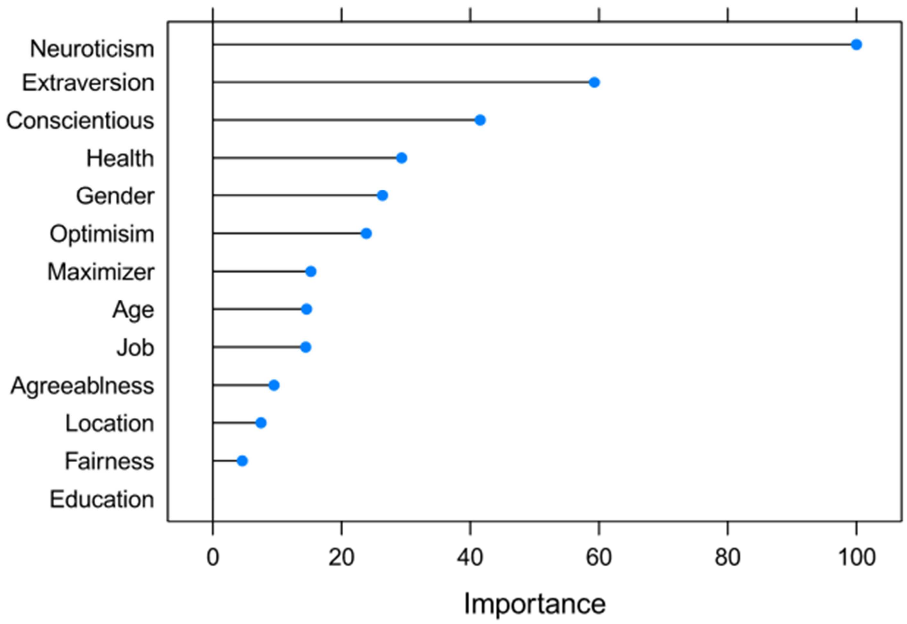


Figure 4.8: Predictor importance in GLM (t-statistic)

The indicated results support previous research identifying neuroticism and extroversion as the most important factors by far (Steel, Schmidt, and Shultz 2008; Hall, Caton, and Weinhardt 2013) followed by conscientiousness and the self-reported healthy lifestyle. Notable is that neither differences in location nor education have a strong impact on the prediction accuracy, contrary to previous literature (Okulicz-kozaryn 2011; Blanchflower and Oswald 2008; Mitchell et al. 2013). For the regression coefficients see Figure 4.9.

22 Also called root-mean-square-deviation.

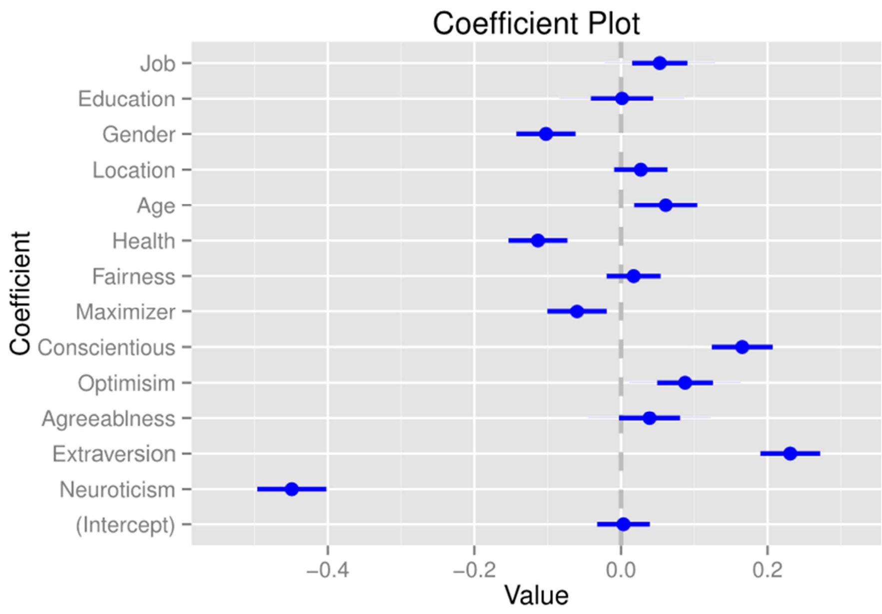


Figure 4.9: GLM Regression coefficients with standard error bars

As seen in Figure 4.8, neuroticism is strongly negatively and extroversion strongly positively correlated with the HFS. Gender is negatively correlated; indicating male participants tend to report lower well-being than female. Education, fairness, location, age and employment situation have no significant influence on well-being ($p > 0.1$). Notable is the comparably strong negative correlation of the personally perceived health. The healthier the participant judges himself to be, the lower is the measured well-being index. The origins of this result are unknown and not discovered in subsequent analyses. The coefficients of the GLM are listed in Table 4.8.

In order to test for possible interactions, the GLM was fitted with linear interaction terms. The non-cross-validated fit has an RMSE of 0.55 (compared to the GLM without interactions: *RMSE = 0.66*) with a significant, positive interaction term for optimism * age ($p < 0.05$). However, if the GLM with interactions is 10 times repeated 10-fold cross-validated, the accuracy drops to RMSE = 0.83. Consequently, the interaction terms do not explain structural variance, but rather over-fit the data.

The results are of the general well-being prediction problem with the averaged well-being index per person as the dependent variable. The results displayed in Figure 4.8 indicate that no linear dependency exists between the 13 predictor variables and the dependent variable, which is the normalized SD between the four HFS measures per participant. All predictors are not significant ($p > 0.05$) and the overall 10 times repeated 10-fold cross-validated model explains less than 1% of the variance within the participants HFS SD (*RMSE = 0.999*).

Table 4.8: GLM coefficients with no preprocessing, 10-fold 10 times repeated cross-validated

	Estimate	Std. Error	t	Pr(>| t |)
(Intercept)	-.003	.052	-.059	.952
Neuroticism	.039	.069	.576	.565
Extroversion	-.059	.059	-1.003	.316
Agreeableness	.044	.060	.732	.464
Openness	.004	.056	.075	.940
Conscientiousness	-.044	.059	-.738	.460
MS Scale	.078	.059	1.325	.186
Fairness	-.064	.054	-1.194	.233
Health	-.020	.059	-.341	.733
Age	-.053	.063	-.852	.394
Location	.089	.053	1.672	.095
Gender	.002	.059	.042	.966
Education	-.011	.062	-.182	.855
Job	-.007	.055	-.133	.894

A similar analysis has been conducted on the slope of each participant's well-being trajectory. To do so, each participant's four HFS data points were separately fitted with a linear regression. The regression coefficient indicating the slope was then normalized and used as dependent variable in the GLM. However, the resulting GLM does not explain any variance between the participants well-being slope upon the 13 predictor variables (*RMSE > 1*). None of the predictors had a significant influence ($p > 0.05$).

4.3.2 Summary and Comparison

RQ 2.1 addresses the ability of well-being data to be used for prediction of participants' well-being baseline and the corresponding well-being trajectory upon the psychometric and demographic input variables. Different machine learning approaches have been tested. However, the algorithms do not achieve a combined higher accuracy and explained variance than the generalized linear model. Three possible causes would explain the obtained findings: Firstly, the conducted algorithms might not be able to fit the existing structure within the data sufficiently. Secondly, the existing dataset is too small in order to differ between structural variance and noise, so that cross-validation eliminates existing structures. However, the accuracy analysis for smaller subsets does not indicate large accuracy gains by larger samples. And thirdly, the linkages between personality as well as demographics and well-being are fairly linear and consequently well-described by the generalized linear model. These linkages have proven to be quite robust and consistent with literature, and can be taken as a design requirement for further TSR applications. It also supports Chapter 3's proposed TSR extension of micro-level factors, as personality and well-being are strongly correlated.

According to the algorithms performed, neuroticism is the predominant variable, followed by extroversion and conscientiousness, which is in accordance with the existing literature. As a new measure in well-being literature, the maximizer-satisficer scale and the participants' fairness perception, have been tested for influences. The first mentioned is found to provide reasonable contribution to the well-being baseline explanation when analyzed by non-parametric algorithms, since a local U-shaped curve has been found in some analyses. However, it is the recommendation of the study to rely of GLM for further predictive models. Fairness perception did not explain additional variance and should consequently not be considered as relevant in subsequent analyses. The same is true for most of the demographic variables, with the exceptions of gender and age. The participant's education, employment and location did not provide any added value. Whereby, it has to be noted that this study's sample is not sufficiently representative with regards to location.

When applying psychometrics as predictors (namely neuroticism and extraversion, along with others) in a generalized linear model, well-being data has shown its suitability for TSR applications. With a partial positive verification of **RQ 2.1**, the research moves on to iteratively and fully address the question.

4.4 BeWell: Prototyping A Game of You

Building on the previous sections, the proof of concept Facebook app **BeWell: A Game of You** is introduced. The app's key aspect is to calculate repeatedly a user's HFS. With a focus on community management and the various concerns thereof, this section presents a method to calculate individual HFS based on (Huppert and So 2013; John, Donahue, and Kentle 1991). Here, gamification comes in: BeWell POC seeks to encourage participants to provide data necessary for the calculation by applying gamification methods in a Facebook application. Being a web application, BeWell POC additionally takes advantage of cost-efficient and real-time data collection and analysis, amongst other things, as well as mechanisms of participant motivation and incentives for truthful information revelation. Section 4.4.1 discusses the gamification methods employed; Section 4.4.2 focuses on implementation of the artifact.

4.4.1 Iterative Design in Gamified Well-being

The interface is built as a Facebook app; as the most popular social network platform with the most established APIs, Facebook is a prime platform for the inception and engineering of new participatory technologies to access well-being information. Flourishing scores are accessible to participants throughout the game. Individual well-being scores, defined by survey responses to Human Flourishing questions, are the means by which one creates their own well-being map. During registration, participants authorize profile data access rights of demographic information including age, gender, location, and highest level of education. Demographics are central for clustering participants based on common identity markers. When participants are linked with various well-being aspects and common identity markers, clustering of participants based on wider identity aspects than their initial network is enabled. Access to post on the participant's timeline for achievements like level completion is requested as a social reinforcement of rewards, and a participation incentive mechanism. The high-level architecture is detailed in Figure 4.10. BeWell POC was available in English and German.

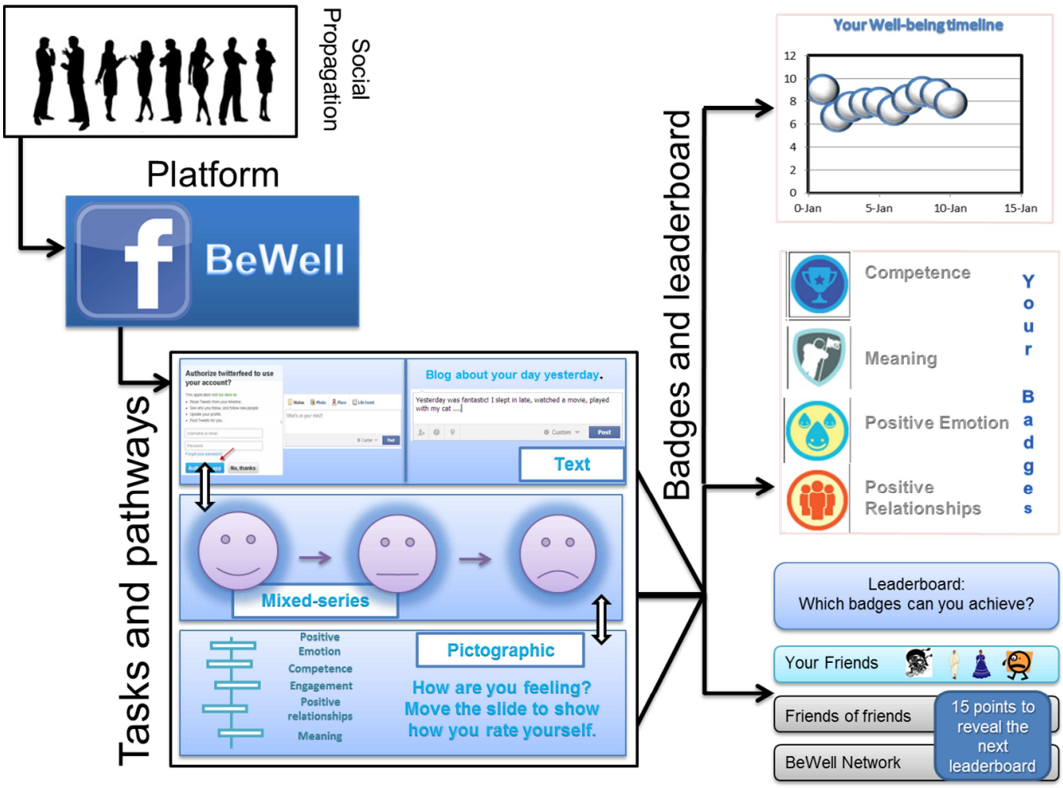


Figure 4.10: BeWell: A Game of You on Facebook component design

Tasks, Missions, and Levels

The interface is accessed in different echelons: a Human Flourishing related question set of tasks; the response mechanism; a portal to view personal game statistics, points, and flourishing score; and a pathway for individual and social comparison. Tasks are the main activity of the game. Tasks are questions based on both on exogenous factors like weather and life events, and nine of ten items of Human Flourishing (competence, meaning, engagement, positive relationships, emotional stability, self-esteem, optimism, resilience, and vitality). These nine items are the game missions. Tasks assigned in groups of either positive functioning or positive characteristics, and are pushed in a reminder format. Each task is offset by a question on positive emotion, the tenth item of Human Flourishing. Positive emotion is named as essential to well-being in SWB as well as PWB, and is therefore a requirement for task completion. After a task series is done, the participant moves to the next flourishing item. Participants who finish all tasks in either of the missions comprising the positive functioning level or the positive characteristics level are rewarded with a level up to either the uncompleted level, or a new treatment group.

Treatments and Pathways

The use of three treatments is a research innovation; previous surveys of well-being are done via a singularly formatted questionnaire or one-shot focus group (Diener 1984b; Veenhoven 2008). However, using multiple treatments is a truth revelation mechanism as it checks the user's reporting of their flourishing level through three different representations. This is an important check due to the introduction of gamification. By hosting a well-being survey in a gamified portal, gamified personas could be induced. By validating users against their own well-being data, the risk of incidental research bias is partially mitigated.

Pictographic representations are the first treatment group. Participants are required to build flourishing related graphics to reveal well-being. Pictographic representations of well-being are mapped to Likert scoring mechanisms based on the depiction of positivity and negativity in the emoticons (Figure 4.11). The scaling is related to Huppert and So's flourishing scale (Huppert and So 2013). Task completion means finalizing the pictograph.

Figure 4.11: A pictographic option of measuring happiness levels

Text analysis is the second treatment. Participants give free-text answers to flourishing questions to complete missions. Text gathered from the responses is analyzed for correlation with the Human Flourishing category being tested. Additional clustering could be completed to search for commonalities in well-being representation between unaffiliated participants, revealing new dimensions of well-being definitions. Text-based responses are manually reviewed. Individuals with high personal assessments of well-being can be expected to use a high amount of positive emotion words, a low to moderate amount of negative emotion words, and words that correspond with positive functioning and positive characteristics. Accordingly, text-based tasks are converted to Human Flourishing scores based on the presence and absence of positivity and negativity in responses. However, the input by participants in the text analysis treatment is below the critical mass needed for an appropriate analysis, and is therefore excluded from this analysis.

The final treatment is a mixed-series between pictographic and text-based representation. The analytics function will read the terms and shapes of the exercise to score well-being. Similar to the text only treatment, additional clustering may reveal unpredicted aspects of well-being commonalities or functions that would otherwise remain hidden. This series allows for a more thorough comparison between both the balance question, and the other treatments. Like the pictographic treatment, task completion requires the completion of the entire exercise.

Point Accumulation

Successful completion of tasks and missions grants points that are redeemed for a variety of rewards (e.g., further access into the social graph, proposing rights for new levels, prizes, gift cards). Points are not the participants' well-being score. Points are granted for not only mission completion, but also propagation efforts. A baseline point bonus is given to participants who propagate to friends. By granting points for introductory propagation, participants are enticed to continue both playing and propagating. Highly propagating participants receive an additional point bundle if threshold levels of participants linked to the gamer participate.

A profile screen grants each participant full access to view their own well-being history, and points comprised of task, mission, and level completion. Point scores and the gaming network's aggregated well-being scores are also accessible in the profile (Figure 4.12). Beginning with their personal network, participants unlock the aggregate scores of further extensions of the games social graph with level completion. This use of personal versus social comparison is in place as a participation incentive, as social comparison is only accessible with point accumulation.

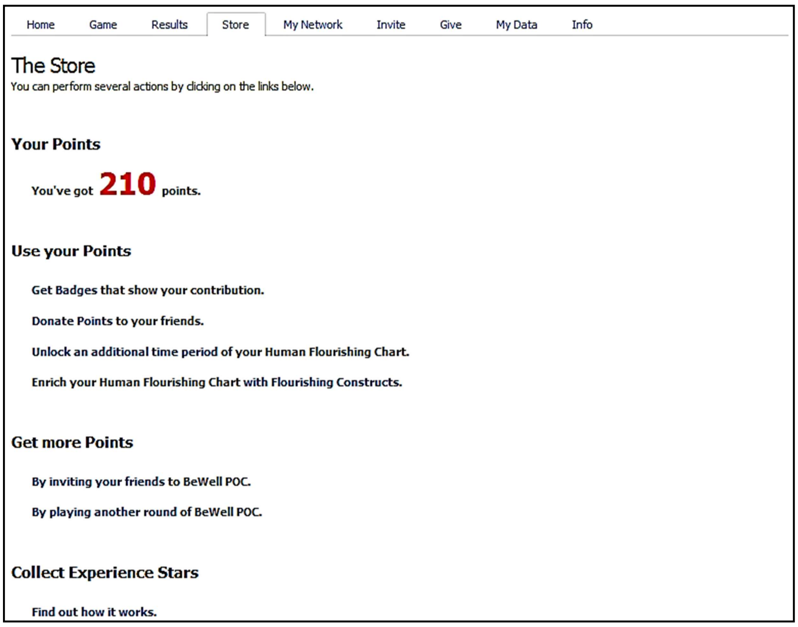


Figure 4.12: The tab "Store" with optional display items

Propagation

The app tracks propagation mechanisms of the game (the way in which participants recommend or advertise the game) and participation in the game (an individual's usage). Tracking propagation helps define online relationships; understanding online relationships is necessary when utilizing social comparison as a feedback mechanism. Participation in the game is the way in which participants populate the data map.

From the perspective of a TSR application, understanding group anatomies and social structures not only aids game design, but also provides an additional management context. For example, a participant with a "poor" well-being score may in parallel be socially isolated (e.g. a new employee). Therefore having access to the social graph can help in the implementation of mechanisms to improve well-being or tackle aspects of low well-being. Looking instead at the implementation aspects of the game, understanding how participants draw in their friends, and the factors that motivate them to do so, enables a better understanding of the relevant social channels. This is important, as without properly addressing the ability to reach as many potential participants as possible, the usefulness of TSR and well-being in particular as an indicator for community and institution health is limited.

4.4.2 BeWell Architecture

Figure 4.13 shows its basic architecture and core components, which are described below. Demographic information was procured via Facebook Permission allowances, with a tab in the game to allow for corrections of misleading or wrongly entered data. The **Question Engine** therefore provides the ability to define arbitrary questions for the measurement of well-being. Questions have three types: 1) a Likert scale question: a question text with a slider; 2) free text question; 3) an animated scale: a pictographic implementation of a Likert scale. Similarly, questions fall into the different categories to fulfill different purposes: 1) Human Flourishing, 2) the Big Five Inventory, 3) the Maximizer Scale, and 4) placebo questions. Fairness was found in the previous analysis to have a minimal effect in personal assessment of well-being, and was dropped in the proof of concept iteration.

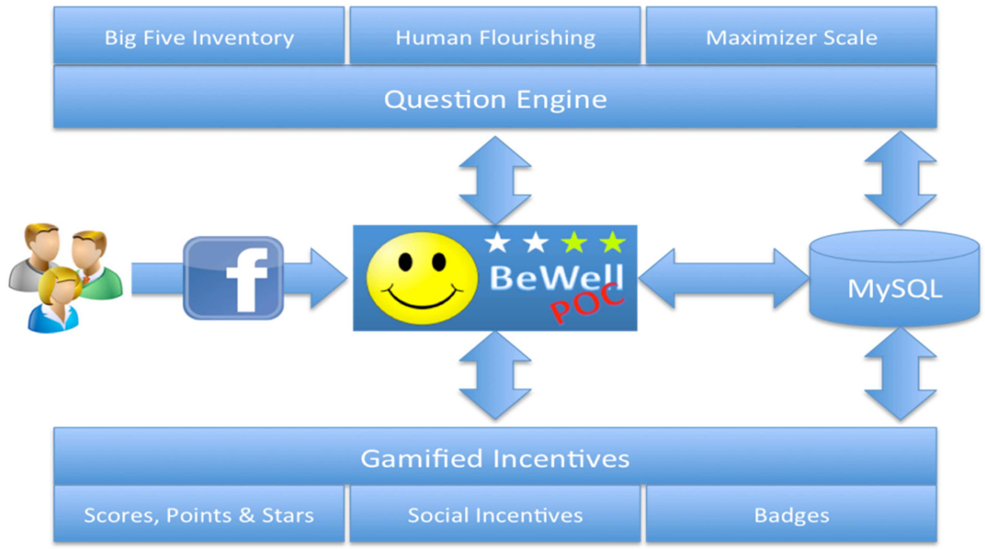


Figure 4.13: BeWell architecture

Game Engine: BeWell POC requires some logic to ensure a smooth data capture process, and to minimize inaccurate data entry. Therefore, participants may only answer questions every eight hours, and in each time period only up to ten questions in order to discourage random clicking. Eight hours was deemed to be suitable time period as it is relatively near in time (closer than a daily interval, for example), and allows for the capture of multiple time periods in a given day (as opposed to half day intervals). Each time period presents participants with randomly drawn questions from the Question Engine.

Gamified Incentives are anchors and features that emerge over time in an attempt to hold the interest of the user, and encourage them to continue answering questions. Three types of incentives are available for users: 1) Scores, Points, and Stars; 2) Social Incentives; and 3) Badges. The types of incentives are explained below.

Scores, Points, and Stars: Key parts of the BeWell POC are the HFS, and allowing the user to track this information. Observing how it changes over time and breaking down its individual components should capture motivate intrinsically. Participants are presented with their HFS graphically (see Figure 4.14 for an example). The graph requires three rounds of questions to be completed before enough data is available (the red line in Figure 4.14). Points are earned by completing tasks in BeWell POC, where the primary tasks are answering questions, and inviting Facebook friends to take part. Points enable a user to unlock the Human Flourishing graph (Figure 4.14), extend it with additional items, and purchase Badges. Experience Stars (as in the logo of Figure 4.13), are earned when a user achieves something, e.g. completes a round of questions, invites friends, unlocks the Human Flourishing graph, buys a badge etc. Experience stars become more embellished with progress and are always visible.

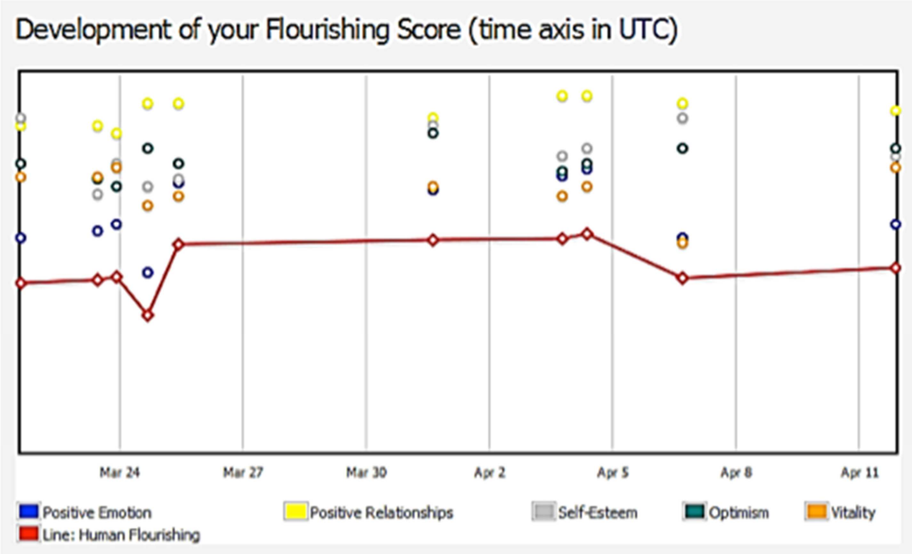


Figure 4.14: Example Human Flourishing score graphic

Social Incentives are constructs that promote social comparison on how well players are progressing, but not on their individual well-being. This is encapsulated by the display of stars, and Badges earned by other players in a user's network. Participants may also send points to their friends, brag about the purchase of items via status posts, and invite friends to take part.

Badges follow the basic principle of trophies that display how far a user has advanced. In an attempt to engage intrinsic motivation (Antin and Churchill 2011; Deterding 2011) badges were designed in the two distinct and leveled flavors "Scientific Advance" and "Better World". They can only be acquired using points earned from answering questions or inviting friends (Figure 4.12). They are incremental (i.e. they can only be purchased in order), and increase in cost. In total, 10 Badges were available (Figure 4.15) and ranged in price from 50–500 points.

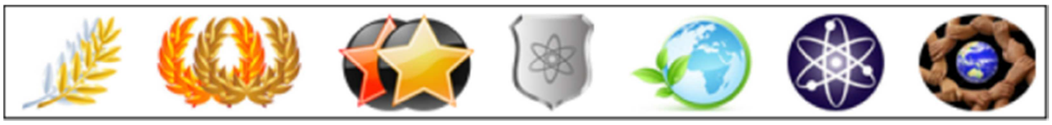


Figure 4.15: Badges available in BeWell

4.4.3 BeWell Pilot Study

The final iteration ran its test phase on Facebook for the period of one month. This version was launched in a gamified environment using the psychometric tests from the previous iterations. The game was propagated through personal networks and was advertised on Facebook via university department websites. The game was offered in both English and German. An additional evaluation user feedback survey was conducted one month after the initial launch with a questionnaire built with the Question Engine.

From the 121 individuals who navigated to the landing page, 37 self-reported to be female and 82 reported to be male. Two individuals did not disclose their gender. 102 participants reported their country of residence to be Germany; eight reported other European countries; and 11 participants are outside of Europe (with seven from the United States being the largest sub-group). Figure 4.16 depicts the distribution of the participants' HFS where n=63, the mean is 44.34, and the SD is 17.44. The distribution resembles that one presented Sections 4.1 and 4.2 with a relative left-shift of around 10%. This is plausibly explained through the fact that **BeWell's** population is tend to be European.

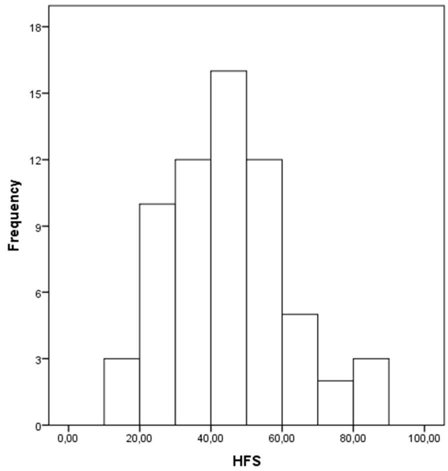


Figure 4.16: HFS histogram of BeWell POC

The analysis also replicates the findings above, namely that neuroticism and extroversion are the two most fundamental predictors of happiness from an individual's baseline personality. Here, correlations are significant at the 1% level corresponding to Extraversion [r(61) = .32, p = .01] and Neuroticism [r(61) = -.39, p = .001]. In this iteration, conscientiousness is also highly significant [*r*(61) = .33, *p* = .007].

Table 4.9: Mean HFS comparison across genders

	N	Mean HFS	Std. Deviation HFS
Male	40	40.06	16.52
Female	22	40.89	19.24
Total	62	44.22	17.55

Men self-report higher flourishing scores (Table 4.9). Due to the overall low participation rate of women, this could be an exceptional case when compared to the results of Sections 4.1 and 4.2. The strength of the deviation between the two genders is in all cases interesting (Figure 4.17). An additional search for explanatory factors regarding higher SDs in the development of Human Flourishing scores was performed. Controlling for demographics, usage activity, and psychological tests no statistically significant explanatory factor was found. This is a mixed result requiring further research.

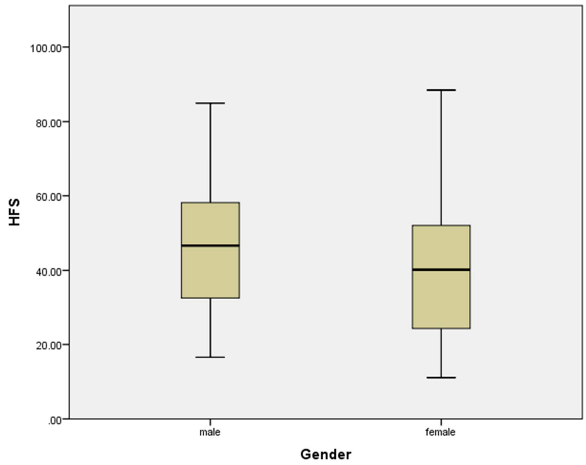


Figure 4.17: Human Flourishing comparison by gender

Design Questions on Flourishing and Gamification

To address how well different gamification structures incentivized continued usage, a questionnaire was built into BeWell POC and activated after one month of data collection. The analysis also revealed limitations of BeWell POC, as well as conflicting results for some incentives. To investigate the irregularities mentioned above further, the data was additionally analyzed for other possible explanatory factors using Spearman's ρ. There are additional correlations significant at the 5% level regarding some questionnaires items. The higher the HFS (and consequently the higher ones extraversion level), the more a user likes “The Point System” [r(28) = .39, p = .034], “Calculation of my Human Flourishing score” [r(27) = .44, p = .016], and “Charting of my Human Flourishing score's development” [r(28) = .41, p = .025]. A higher HFS further correlates to less enjoyment of “Posting Badges to my Facebook timeline” [r(28) = .40, p = .028].

Remarkable is the high number of significant correlations found between the personality trait “Neuroticism” and the incentives. A highly significant negative correlation with neuroticism can be seen for the items “Getting Experience Stars” [r(29)=-.59, p = .0005], “Getting Badges” [r(31) = -.56, p = .0008], “The Point System” [r(29) = -.47, p = .008], and “Comparing my Badges to those of my friends” [r(29) = -.46, p = .01]. A negative correlation with neuroticism still significant at the 5% level can be seen for the items “Comparing my Experience Stars to those of my friends” [r(28) = -.41,p=.024] and “Inviting Friends” [r(30) = -.35, p = .049]. As the scale used in this part of the questionnaire implies that an item is more liked the higher its value, a negative correlation means: The more neuroticism participants report, the more likely they are to dislike these specific incentives, which can have important (and complicated) design implications. The complications stem from the fact that when measuring well-being, neuroticism and extraversion are the strongest predictors (Section 4.3), but the two traits' acceptance of the gamified elements of the survey is in opposition. Gamified elements are attractive and accepted by extroverts and rejected by neurotics. This almost suggests that two game flavors should be developed in order to entice participation from all personality types. While intriguing, element design for neurotics is outside of the scope of this thesis.

The findings indicate that while there is still some work to be completed on the incentive mechanisms, this goal is in fact achievable. Looking at the gamification incentives, one can see that the primary interest of the participants was to calculate and track their HFS, and to investigate their Flourishing items. They predominantly seem to have liked the gamified approach that was taken. Badges and Experience Stars were of lower importance, but still liked. This is not true for the bragging feature (posting of Badges to one's Facebook timeline) which was clearly unused. The social incentives built into BeWell POC were also underutilized, supporting the view that the participants were rather self-contained. Not surprisingly, however, the valuation of the possibility to compare Badges and Experience Stars to friends, as well as to see who is also a user of BeWell POC, is dependent on the actual number of friends playing. This indicator supports the plausibility of the participants' responses regarding the questionnaire. There is an observable rejection of comparative and evaluative incentives through participants with higher neuroticism levels.

4.5 Discussion and Limitations

This chapter proposed a gamified approach to well-being data collection, some potential overlapping decision areas, and challenges of propagation in future TSR applications. It presented a methodology that utilizes attributive predications in order to analyze data obtained in gamified systems for progressive community management, and evaluated the feasibility of acquiring well-being data via online social networks by collecting near to real-time data in a longitudinal rather than cross-sectional manner. The results aided in the realization of BeWell's proof of concept app in that they provide a guideline for the development of future predictive models. BeWell POC was found to responsively track trends in noisy data of personal well-being, continually updating given the collection of new data points, and highlights otherwise hidden attribute-based well-being forecasting.

Importantly, a tiered phase-in of the BeWell concept was implemented. Each iteration expands the initial scope in length and questions utilized. The pilot was the first instance of Human Flourishing being utilized in an online format. All questions of the flourishing survey were mandatory, and optional demographic data of gender, age, place of residence, and highest completed education level were optional. The ten questions were positioned online for one week and initially propagated on Facebook. Questionnaires were available in English and German.

The next two iterations of feasibility surveys were propagated in online social networks in order to validate if attribute-based prediction can be used in conjunction with the measurement of well-being. Surveys were administered online once per week for four weeks on Wednesdays, in order to control for variance in weekly activities, such as subjective preferences for weekends. Ten identical questions covering varying aspects of Human Flourishing were posed to facilitate prediction of said dimension. Demographic questions, the 44-item scale Big Five Inventory personality test, the Maximizer/Satisficer scale test, and a fairness scale (John, Donahue, and Kentle 1991; B. Schwartz et al. 2002; Schmitt and Do 1999) were added as potential predictor attributes. Each psychometric instrument was administered for one week only to test prediction abilities of well-being based on pre-existing personality traits.

The feasibility studies confirmed the ability of psychometric properties to predict levels of well-being (**RQ 1.1**). Two factors of the Big Five Inventory, namely neuroticism and extroversion, are observed to have the highest predictive value, especially when analyzed with a general linear model. The findings also reveal interesting discrepancies with previous work; namely, that conscientiousness is in fact a significant baseline personality factor, and that the maximizer-satisficer could in fact be U-shaped. The outcomes from this analysis illustrate the ability to predict well-being in a future TSR application. These results support the creation of attribute based tracking for the establishment of baseline well-being expectations. Using these attributes, well-being baseline assessments are creatable for use to predict future well-being values. Manifestations of the absence of well-being or a change from its expected level are predictable when plotted, thus facilitating evaluation and stakeholder discussions. The vision of gamified well-being revolves around the use of smart devices, in the context of a familiar setting (Facebook), which should facilitate the construction of a progressive community portfolio: a stakeholder feedback loop of community well-being and overall satisfaction.

Regarding incentives, improvements are possible. An observed drop in participation after four iterations was visible in both the feasibility studies and the proof of concept app. For active participants, a new version could relax the prerequisite to bring up all ten Human Flourishing related questions per round. Instead, the period considered for the calculation of the current HFS could be extended and span answers from different rounds. This way, e.g. five flourishing-related questions could be generated per round if the last round was not too long ago. With gamification now shown to be functional, it would be possible to push the rather limited range of questions further, moving into the direction of a “Game Engine” for different sorts and complexities of tasks. The bragging feature was left unused. There is no reason to keep it in future versions. A method to opt-out from comparative and evaluative incentives is also required, as a many participants disliked them. One could imagine a setting that to hiding the respective links in the tab “Store”; disabling the assignment of Experience Stars; and disabling the display of Badges and Experience Stars.

Self-produced text solely for the purpose of the gamified environment does not incentivize participants to sharing. However, further research is needed to confirm if Facebook will continue to be a viable platform. Potential issues include decreasing popularity, self-representation in online social networks, and other issues of truthful reporting (**RQ 2.4**). Finally, distribution of the three iterations suffers from a CMB (Podsakoff et al. 2003; Conway and Lance 2010); namely, the directed nature of participation invitation lends itself to reference and self-selection biases, thus the results reported here must be interpreted with caution.

BeWell POC collected additional data that has not been detailed in this chapter. Examples include analyzing of additional usage tracking data and testing for possible significant correlations between the placebo and ten Flourishing Questions. Also collected in every iteration were general comments and feedback. This anecdotally suggested that a major participation barrier is the time required to play the game. This could be the contributing factor to the observable drop in participation after four rounds (**RQ 2.2**). General next steps are to integrate the findings presented in the above section into new versions of BeWell POC. A serendipitous finding is the valuable service that the notifications feature provided. User reaction was clearly tracked and reported, and some participants became “chart unlockers” and long-term players as a direct result. Future versions should build on that, e.g. by providing user-customizable notifications (email is also a possible channel) with a sensibly preset interval.

4.5.1 On Serious Games for Well-being Assessment

The final iteration addresses **RQ2.2** in its full breadth, and partially fulfils **RQ 2.1**. This iteration was created as the proof-of-concept application, integrating and extending the features introduced in the first two research and design phases. Implementing and assessing the well-being of a community or institution via popularly propagated social gaming is a novel person-to-person mechanism in computational social science. This work establishes that serious games are a suitable method for the extraction of well-being data, but suffer from participant fatigue. As such, this thesis moves forward with text analytics as an extraction method (**RQ 2.3**).

Rewards are layered upon existing activity, with flourishing items as tasks, and entire constructs as missions to be completed, allowing point accumulation and level achievement. The ability to chart oneself, the gaming community, and earn points-based prizes serves as rewards and incentives for continued participation and propagation. Propagation is further encouraged via social action - reaction prompts on open profiles and direct invitation. Social interaction creates an incentive to participate, and reciprocate.

Eliciting well-being via a person-to-person game induces the experience of personal perception and social comparison within an online community. Given the strong replications of the relationships between personality and well-being, it can be rejected that participants are using ‘gamified’ personas in their responses to the gamified environment. In gamifying, participants are incentivized to reveal their personal estimates and are encouraged to propagate the game further across their social graph. This is a partial response to **RQ 2.4**.

4.6 Conclusion

Online gathered and popularly sourced well-being information is ripe for adaptation into TSR. By utilizing such a multi-faceted picture of the individual, BeWell encourages communities to proactively manage the components causing agency loss (e.g. cheating, lack of transparency, ill-health) as a form of adaptive people management. Such an elastic measure can be repurposed as both a diagnostic and predicative model for diverse participation-based movements and institutions when populated with well-being data. This supports the aims of TSR well. The next steps are mapping well-being to communities, regions, and institutions to illustrate policy effectiveness and enhance participative debates. Through the observation of a statistical decrease in well-being, participatory approaches could be a reactive measure as a means to reengage constituents, and engage new participants throughout the community. Gamified well-being measurement has proven to be a reliable and valid data population method for progressive community management.

However, BeWell’s dependency on engagement and propagation of the crowd and community are a suboptimal basis for the development, measurement, and management of social indicators such as those proposed in Chapter 3. The chances that failing interest curtains participation cannot be underestimated. Also, the self-selection bias of those who participate in a non-mandatory measurement tool can influence results in an undesirable way. Estimating the reach of a representative community is also difficult in this case. Promising directions for the measurement of well-being in the efforts towards progressive community management are those which are unobtrusive, or that have little to no observation effects, and that mitigate self-selection bias and participation dependencies by being previously well-established in a community. Whilst BeWell and its proof of concept Facebook app satisfactorily addressed **RQ 2.2** and partially addressed **RQ 2.4**, further investigation of alternative mechanisms for a TSR application, namely text analytics, is pursued in the coming chapters in accordance with **RQ 2.1**.

Chapter V Online Well-being: An Applied Social Observatory

"It's representative of the moment we're having; We talk in hashtags. It's how we share information right now."

Brett Hyman as quoted by (Meltzer 2014)

With social media, political parties bring their message to the public faster, positing on recent events before the interaction and interpretation of local or national media. Putting issues onto the public stage they can directly interact with voters, supporters or residents of their election districts, thereby acting locally as well as nationwide. As such, political discourse is similar to the changeover in the serivitized, digital economy. However, what is currently missing is a valid measurement system (e.g., a TSR application) that allows insights into the way policies and current political discourses are being received and the impact thereof. Such a system in conjunction with data from public information sources could assist social researchers and decision makers with the analysis, development, implementation and tuning of policies. Specifically text gained from online sources can be spliced for context and content, compared, and measured for sentiment and conceptual domains as a means of well-being assessment. Sentiment-based artefacts using publicly available data thus promises unprecedented access into the expectation of arising changes in well-being ex-ante, and the totality of effect of incidents ex-post. As such, text and sentiment analysis is well-poised to support a TSR application.

A new approach in information-driven TSR is the utilization of the measurement of public discourse and sentiment levels for "mood management" to gather prompt, direct feedback on arising changes within affected communities. A requirement for this is that information can be unobtrusively gathered to assess public sentiment (Section 5.1). Given the possibilities and enormous user base, the social network platform Facebook is an interesting test bed. Facebook empowers users to publish opinions and causes, and publicize and document activities to solicit ones work, products, or beliefs, and is a ubiquitous part of digitalized lives. Expressed there are not only thoughts and opinions but (latent) feelings and expressions of well-being. This chapter presents an extraction method called the Social Observatory: *an unobtrusive, low latency, multi-resolution framework for the observation, analysis and modelling of digital societies in action*. With a Social Observatory, this research realizes an automated framework that facilitates, reviews, and assesses specific aspects of online communities (e.g., well-being) using qualitative and quantitative methods (Sections 5.2 and 5.4) as a facilitator of the aims and goals of TSR. The research objective is a framework that empowers interdisciplinary researchers with the tools to facilitate the extraction and understanding of phenomena within social media platforms, as well as the communities they represent.

This chapter presents a prototype implementation and case study analyzing public political dialogue of German federal politicians on Facebook (Section 5.3). The dataset is comprised of all politicians with a Facebook presence from the five German federal parties: the Christian Democratic Union (CDU/CSU), the Social Democrats (SPD), the Free Democrats (FDP), the Green Party (Grüne), and The Left Party (Die Linke). 52,833 posts and 267,835 comments are analyzed, creating a composite index of overall public sentiment and well-being, and the latent conceptual themes supporting this. Our case study demonstrates the observation of communities at various resolutions; "zooming" in on specific subsets or communities as a whole to view various granularities. The results of the case study illustrate the ability to observe published sentiment and public dialogue as well as the difficulties associated with established methods within the field of sentiment analysis and topic retrieval within short informal text.

This chapter extends two sources: the journal article (Caton, Hall, and Weinhardt, forthcoming) as well as a working paper presented at the Karlsruhe Service Summit Workshop (Caton et al. 2015).

5.1 Big Data Challenges in the Social Sciences

The vision of a Social Observatory is a low latency method for the observation and measurement of social indicators. It is a computer-mediated research method at the intersection of computer science and the social sciences. The term Social Observatory is used in its original context (Lasswell 1967; Hackenberg 1970); the framework is the archetypal formalization of interdisciplinary approaches in computational social science. The essence of a Social Observatory is characterized by (Lasswell 1967, 1) as follows:

> *"The computer revolution has suddenly removed age-old limitations on the processing of information [...] But the social sciences are data starved [...] One reason for it is reluctance to commit funds to long-term projects; another [...] is the hope for achieving quick success by 'new theoretical breakthroughs' [...] It is as though we were astronomers who were supposed to draw celestial designs and to neglect our telescopes. The social sciences have been denied social observatories and told to get on with dreams."*

This is also in line with the approach of the American National Science Foundation's call for a network of Social Observatories:

> *"Needed is a new national framework, or platform, for social, behavioral and economic research that is both scalable and flexible; that permits new questions to be addressed; that allows for rapid response and adaptation to local shocks […]; and that facilitates understanding local manifestations of national phenomena such as economic volatility."*[23]

Today, the notion of a Social Observatory lends itself towards social media platforms, as digital mediators of social exchange, discourse and representation, as well as to the multi-layered approach introduced with TSR in Chapter 3. This, as demonstrated by the COSMOS project (Burnap et al. 2014; Housley et al. 2014; Procter et al. 2013), becomes especially valuable when combined with government data streams. However, empowering social scientists to access data from social media platforms (even in the singular) is non-trivial.

Figure 5.1, illustrates a general architecture of a modern Social Observatory entailing three processes; namely 1) Data Acquisition; 2) Data Analysis; and 3) Interpretation. Whilst it is apparent that a Social Observatory captures multiple sources of data, currently few scientific papers or services report this ability in a way easily replicable by social scientists (Cioffi-Revilla 2014). This is despite prevalent availability of APIs, and an almost endless supply of papers and studies that focus on specific platforms (Russell 2013).

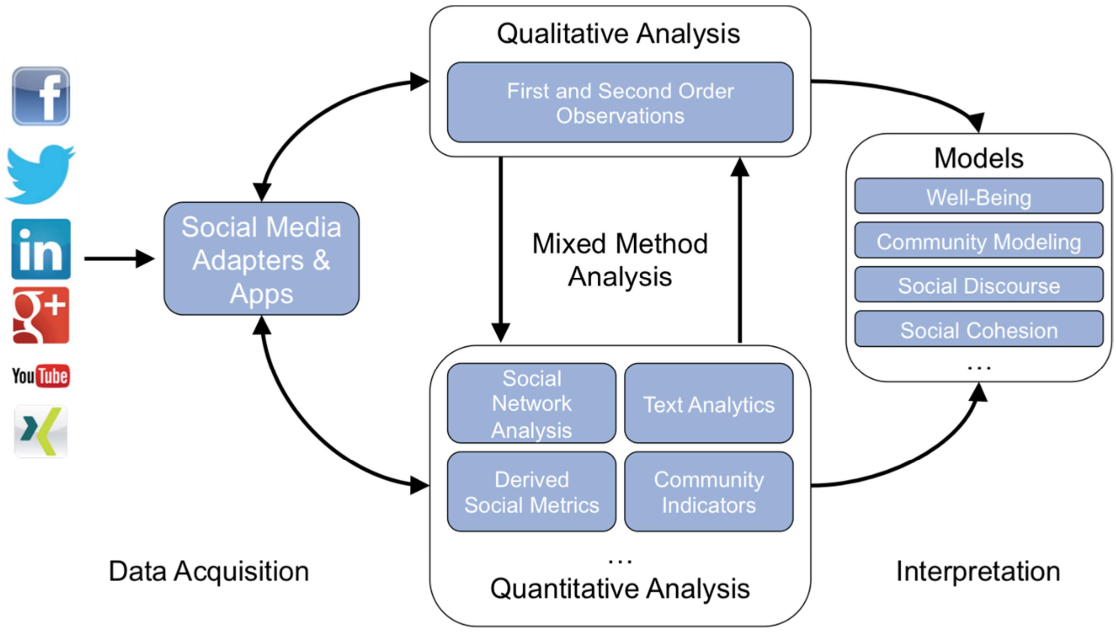


Figure 5.1: A General architecture for a Social Observatory

[23] http://www.socialobservatories.org/vision. Last Accessed: 01 October 2013.

Data Acquisition is well supported by most social media platforms via REST or streaming APIs, which are underpinned by lightweight data interchange formats like JSON. User authentication and access authorization is handled by technologies such as OAuth. There are also an ever-increasing number of software libraries available, reducing the implementation effort to extract data.

The challenges instead lie in data volume, velocity, and variety, access rights, and cross-platform differences in curating data. The big data aspects of social media data are well known: producing 2,200 tweets (at around 58kilobytes each) per second, Twitter is a clear demonstrator of data volume and velocity. Variety is best shown using a Facebook post as an example: version 1 of Facebook's Graph API contained at least 15 categories for a user post and this discounts other social actions like tagging, commenting, poking etc., as well as the diverse content range of a Facebook user's profile. Lastly, the method of data curation is not without its ambivalence. Twitter data curation tends to be proactive; by accessing future tweets that fulfil a specific set of user-driven attributes (e.g., hashtags or geolocation). Facebook is retrospective; given a Facebook entity (e.g. a person, or page), one can access their posts, profile, likes etc. From the perspective of analyzing social data, this subtle difference significantly alters the effort and planning needed to curate a data set (González-Bailón et al. 2014). The technical challenges also differ significantly from receiving a continuous stream of data (i.e., tweets) vs. Facebook's paginated results. The latter incites large numbers of API calls, which are not limitless. On a side note, the validity period of an access token is also not infinite and must be refreshed periodically.

(Mixed Method) Analysis as illustrated in Figure 5.1, is inherently iterative and interdisciplinary. Foreseeable is repeated interaction with the social media adapters and apps. Whilst approaches from computer science and computational social science are becoming more prevalent, the question of research methodology is often a poignant discussion point and challenge that cannot be overlooked. Computer scientists and social scientists speak very different languages. Therefore, the realization of a Social Observatory needs to accommodate a vast array of (interdisciplinary) methodological approaches.

Irrespective of methodology, an important feature of a Social Observatory is the ability to view a community at a variety of resolutions; starting from an individual micro layer, and progressively zooming out via ego-centric networks, social groups, communities, and demographic (sub) groups, up to the macro layer: community. This ability is of significant importance for understanding a community as a whole; different granularities present differentiated views of the setting. **Interpretation** is hence domain specific in nature, and should be decided according to the proposed research questions. The architecture supports both inductive and deductive research.

Necessary to address at this point are the ethical boundaries of an unobtrusive approach to Big Data analyses of social data. Both Twitter and Facebook have terms and conditions allowing for the anonymized assessment of data which the use has indicated to be public. Specifically Facebook has argued that this is tantamount to informed consent,[24] and this is a common position across social media platforms. This study agrees that when information is placed in public fora and domains, it is subject to public review. This is in line with the ethical guidelines put forth by the Association on Internet Researchers (Markham and Buchanan 2012). In the case of obtrusive design (i.e., greedy apps), informed consent must continue to be in place as the standards of human subject research demand. A further ethical (and security) concern is that the provide architecture can also be used irresponsibly. In the case of public-facing data, this is of a lesser concern. Obtrusively-designed architectures still require user consent (e.g., downloading an app), as such research works are neither the work of hacking nor 'Trojan horses,' thus guaranteeing a moderately informed subject base.

5.2 Social Networks as a Proxy for Communal Well-being

For the past few decades researchers have investigated the interaction of technology, online communities, and individuals' perception within it (Larsson et al. 2005). Similarly, text analytics for measuring social impact is an emerging topic but has not received much attention despite its long-standing recognition (Pennebaker, Mehl, and Niederhoffer 2003; Housley et al. 2014). This research gap presents a novel place for computer science, text and sentiment analysis, and policy jurisdictions to meet. Whereas many of the commonly applied methods in community analysis like judging communal sentiment, assessing strength and weakness of ties, or willingness to participate and/or exchange in a given context is a task easily done manually, manual approaches do not scale. Moreover, it has been established that sentiment and conversation styles differ across platforms (Davenport et al. 2014; Lin and Qiu 2013), though the available tools do not match this research need. The (social) scientist lacks the necessary systems, tools, and competencies to leverage computational approaches. A new approach in the area of information-driven institutional management is found in computational social science (Cioffi-Revilla 2014).

Computational social science (Cioffi-Revilla 2010; Cioffi-Revilla 2014) facilitates investigation of the interaction of technology, online communities, and individuals' perception within it to a previously unmanaged scale (Savage and Burrows 2007; Burrows and Savage 2014; Tinati et al. 2014; Taylor, Schroeder, and Meyer 2014). Text analytics as a mechanism for measuring social impact is becoming ever more validated as a proxy for social phenomena (Mckelvey 2013; Housley et al. 2014; Böcking, Hall, and Schneider 2015). Such a research domain is complementary to the aims of a Social Observatory, where the differences are that

[24] http://newsroom.fb.com/news/2012/05/enhancing-transparency-in-our-data-use-policy/. Last Accessed: 23 May 2012.

computational social science is an entire research domain and a Social Observatory is a framework to enable research thereof. Specific to the assessment of public sentiment, Twitter-based studies are plentiful and address a variety of computational social science research questions. Off the shelf Facebook tools are less well-addressed. Several author have addressed the creation of frameworks for supporting Twitter studies (Stieglitz and Dang-Xuan 2012; Pak and Paroubek 2010; Burnap et al. 2014; Housley et al. 2014). These lack however the corresponding technical infrastructure that allows future researchers to create new, build on or replicate the studies. The closest in reach to a Social Observatory are those where the infrastructure is both open-source and requires minimal knowledge of computational infrastructure in order to be accessed (Burnap et al. 2014), or the tools are of a plug and play nature (McCallum 2002; Kivelä and Lyytinen 2004).

5.2.1 Studies in Online Social Media

In Twitter the use of positive and negative, or positive, negative, and neutral classifications of individual tweets as opposed to more contextual sentiment is a common method (Pak and Paroubek 2010; Burnap and Williams 2014); this is likely due to the shortness of individual tweets. A foundational paper from (Go, Bhayani, and Huang 2009) looked at the classification of Twitter sentiment from the commercial perspective, identifying positive and negative tweets based on query terms of emoticons. (Kouloumpis, Wilson, and Moore 2011) found that intensifiers are most useful in the automated detection of sentiment in tweets. This study found that part-of-speech features are not necessarily useful in automated sentiment detection. A study by (O'Connor et al. 2010) applied positive and negative sentiment scoring to the 2008 presidential elections of the United States and found the method can be used to supplement consumer confidence polls.

Key contribution differences are the observation viewpoint and elicitation of points of reference. Many studies observe the Twitter landscape at a macro level, whereas a Social Observatory facilitates micro, meso and macro observations in accordance with the layered approach set up in Chapter 3. Specifically the micro-level is difficult to realize with Twitter due to the brevity of individual posts. (O'Connor et al. 2010; Calvo and Mello 2010; Hampton et al. 2011) demonstrated the predictive power of self-reported interests in social profiles and the observation of social practices. Whilst the scientific value of such work is significant, they are isolated investigations. For the purposes of TSR applications, they give insights into well-grounded research processes rather than assisting in the construction of a general approach. Similarly, (Mitchell et al. 2013) investigates a macro-scale dataset of happiness, urbanization and obesity correlates, but does not create a generalizable model for wide-scale usage. (Allen et al. 2014; Jaho, Karaliopoulos, and Stavrakakis 2011) investigated how content traversed social graphs, and explored opportunistic mechanisms for the dissemination of content via social structures. A focus of their work was mechanisms for community detection, and subsequent analysis of social structures for observing information paths through social networks. However, the emphasis is on the support of users in identification of content relevant for specific decision making processes, and methods to facilitate the transfer of information via and within social structures, as opposed to analyzing the communities themselves. Finally, Facebook researchers have investigated if positive and negative well-being are contagious; and indeed, the expression of well-being is contagious (Kramer, Guillory, and Hancock 2014). It must be noted that this study actively altered the emotional valence of the study participants' timelines to establish its findings. This thesis attempts to establish emotional valance and trends even more unobtrusively in order to not inadvertently impact individual's well-being.

5.2.2 Related Online Social Media Studies on German Politicians

The study of (Tumasjan et al. 2010) concentrates on the application of Linguistic Inquiry and Word Count to text gained from German politicians' twitter handles in advance of the 2009 elections (Tausczik and Pennebaker 2010; Pennebaker et al. 2007). Their analysis has several distinct differences, elucidated here. This research uses the German dictionary database provided by LIWC2007 (Wolf et al. 2008) for the analysis of online political behavior and discourse, rather than translating to English for analysis to retain the original intention of the writer as closely as possible. The focus of this observation period is the election period of 1 September, 2012 through 31 October, 2013, enabling longitudinal analysis as opposed to a cross-sectional analysis. This supports the study of well-being in a community more fully. Whereas Tumasjan and colleagues review selected LIWC categories, this study considers all German dictionary categories and established psycholinguistic profiles. Finally, the aim of the study is a diagnostic analysis of political messaging on online social media. It is not a prediction task.

5.3 Implementation: a Facebook Social Observatory Adapter

The first step towards a Social Observatory focuses on a Facebook social adapter for several reasons. Firstly, Facebook lends itself to the case study, especially due to the large number of "open" Facebook entities; where community and personal pages are a prime example. Secondly, when extracting data from Facebook, the researcher receives near complete datasets. Finally, there is lack of general-purpose Facebook data acquisition tools available, which is a current research gap. Those that are available tend to rely either on crawling techniques, which cannot fully acquire paginated Facebook data, or data extraction via the Graph API that typically focus on the logged-in user or do not return data in full. Whilst such approaches are useful, especially in classroom settings, they do not provide mechanisms to curate research worthy datasets. This chapter presents a general and extensible Facebook data acquisition and analysis tool: FBWatch.[25]

The objective is simple: an interface-based tool allowing social as well as computational scientists to access complete Facebook profiles irrespective of programming ability or data size, as no such tool is available. In extracting data from Facebook, the researcher first needs to define what is accessed: an entity that has a unique Facebook identifier.[26] FBWatch is implemented such that it can access any Facebook entity that is public, or for which it has received user permissions.

FBWatch is implemented using the Ruby on Rails framework, and consists of five top-level components and modules:

1) A Sync module responsible for fetching data from Facebook. It executes Graph API calls, converts graph data to the internal data structures and stores it in the database;

2) Metrics are the analysis components of FBWatch and responsible for analyzing fetched data. They contain parameters used for case studies and data structures for storing results. A metric can therefore be any result of an analysis (exemplified in Section 5.4);

3) Tasks, which are an abstraction for running Sync and Metric jobs as background processes;

4) A relational database for storing Facebook resource data, and running more complex queries regarding connections between Facebook entities. Any SQL-Database can be used provided that it supports UTF-8 encoding, as this is needed for handling foreign languages;

5) A web front-end as an access point and controller for FBWatch. Here the user can request the retrieval of new Facebook entities, refresh previously fetched entities, group entities together for comparative analysis, execute metric calculations, visualize metrics as well as the social network of individual or grouped entities, and download datasets for use in third party analysis tools.

[25] Accessible via github: https://github.com/luksurious/fbwatch-ruby.git. Last Accessed: 04 June 2014.
[26] Note resource and entity are used interchangeably.

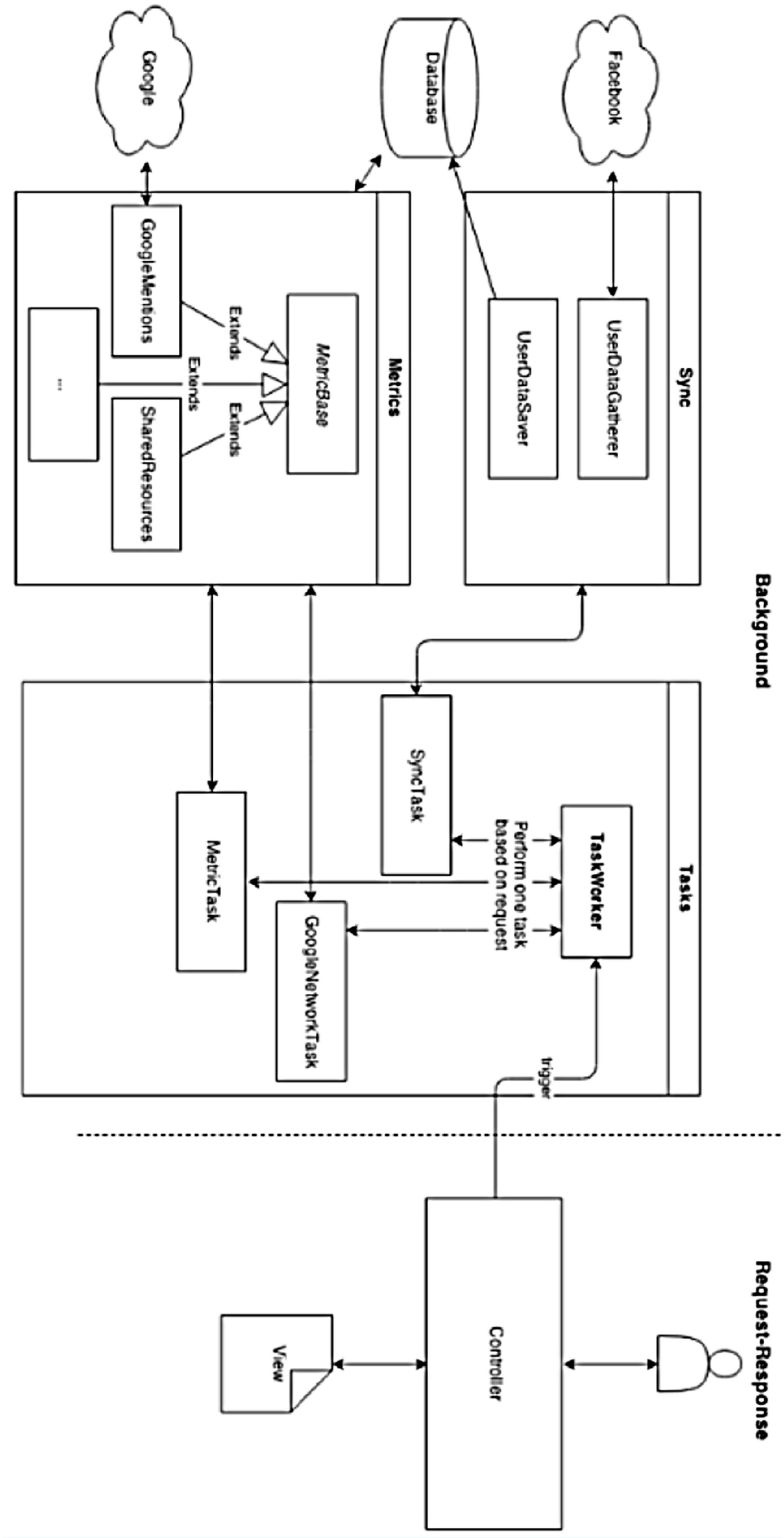


Figure 5.2: Workflow illustrating the steps to acquire, analyses, and interpret Facebook

Figure 5.2 shows the architecture of FBWatch, and highlights a typical request involving either the data fetching, or the metrics calculation. Upon a request, the controller triggers a background worker class and returns an appropriate view to the user who is notified that a task was started. The worker then performs one of two tasks, depending on whether Facebook data is to be retrieved, or retrieved data is to be analyzed.

The first step in the process flow the user providing the Facebook URL of one or more entities of interest, which are parsed for their username or Facebook ID. To synchronize the data of Facebook resources, a background sync task is started by FBWatch. The user can check the status and progress of the task, as required. Depending on the size and number of entities, synchronization can take several hours, and can also encounter several errors that need to be handled manually. Once synchronization has successfully completed, this will be visible and the user informed of how many feed entries have been retrieved. If errors were encountered that could not be handled this will also be displayed.

To access data, Koala, a lightweight and flexible Ruby library for Facebook, is used. It provides a simple user interface to the Graph API and the Facebook Query Language. As the Graph API returns the data in JSON format, Koala automatically parses the resulting string and converts it into the appropriate data structure using Arrays and Hashes and aligns the primitive data types into Ruby's data types. Furthermore, the library supports the use of the OAuth protocol to authenticate within Facebook through the use of the OmniAuth Ruby library. A valid, i.e. Facebook authenticated, instance of Koala is generated on a per-session basis and stored in the session context. At this time this is also the only real authentication the application performs directly. To mitigate exposing all data fetched by FBWatch, HTTP authentication is enforced on the server.

Synchronizing a Facebook resource is done in a two-step process. First, any basic information of that resource is pulled by calling the Graph API link facebook-id.[27] Basic information contains the information visible at the top of a Facebook page and in the about section, like first and last names, website, the number of likes etc. Second, the actual feed data is retrieved.

This is not trivial. First of all, not all data will and can be received at once, as Facebook limits the number of results per query; 25 per default. Increasing this limit drastically reduces the number of Graph API calls, and thus, speeds up the data gathering process. By default FBWatch uses a limit of 900, increasing speed and managing scalability. Facebook also only returns a subset of the comments and likes of a feed item; four by default. The resulting data contains a paging feature, similar to the one of the feed itself in a single feed item. Comment and like arrays have to be fetched using multiple API calls, dramatically increasing runtime. The **UserDataGatherer** module automatically navigates the paging system until it receives an empty data array. FBWatch also stores the link representing the first response from Facebook.

[27] The corresponding command is /<facebook-id>/feed.

This allows FBWatch to easily update a resource at some point in the future. If, however, a problem occurs, the last feed query is stored to enable the future continuation of a sync task.

The second part of the Sync module stores fetched data via the **UserDataSaver**. Aside from transforming Facebook JSON into internal data models, data entry needs to be optimized such that it scales. In order to decrease runtime, multiple INSERT and UPDATE statements are grouped into transactions. However, not all statements can be executed in one transaction due to interdependencies between data models. Thus, saving the data in the correct order is important. In order to take into account all possible dependencies, four transactions are used:

1) Resources and their basic data are updated as well as all new Facebook entities that posted or interacted on the feed at the root level,

2) Feed entries,

3) Resources which interacted at a lower level, i.e. with a comment, like or tag, and

4) The comments, likes and tags.

Once an entity has been fetched, it can at any time be resynchronized to retrieve any new feed items and their properties or continue to fetch all historic data if the synchronization was not successfully completed before. If a resource is no longer available on Facebook or no longer relevant for the analysis it also can be disabled or removed. Apart from the ability to traverse Facebook data automatically using the provided paging mechanism, the other main feature of the UserDataGatherer is error handling. The Facebook API is not reliable all the time, and is badly documented. Therefore, flexible error handling is required. The most pertinent hurdle is a limit to the amount of calls a single application can execute for a given access token in a certain time frame from the same IP address. While it is not officially documented, as per Facebook, apps tend to be limited to 600 calls every 10 minutes. For large resources, this limit is hit multiple times. FBWatch handles this by pausing the sync task, and retrying periodically (every five minutes) to resume it. This can require up to 30 minutes. FBWatch also handles when a resource cannot be queried, be it that it was deleted or disabled, when a username has been changed, and other miscellaneous errors.

5.3.1 Data Model

The data models representing social network data is loosely based on the Facebook Graph API format.[28] A resource model corresponds to one Facebook entity but also constitutes the most important object in FBWatch. All overlapping properties of the different types of Facebook

[28] https://developers.facebook.com/docs/graph-api. Last Accessed: 10 June 2014.

resources are saved in this data model: the free text name, the unique Facebook ID, the unique username and the full link to the resource on the Facebook system. Additional data relevant for the application is saved in this data model as well: a flag indicating whether or not a resource is active, i.e. if it should be synchronized, and the date of the last synchronization.

Other information returned by Facebook differs greatly for different entity types and is thus stored as an array of key-value pairs. Here, information such as the number of likes for pages, a website URL or the first and last names of real users, their gender and (given or Facebook) email address is represented. Furthermore, configuration data of the application is stored: information of the last synchronization so that it can be resumed more easily and no duplicates are retrieved. The value of stores the URL of the first link of the paging feature of the first feed page, i.e. where at the moment of synchronization newer data would be available. A property is called ‘last link’ stores the link to the last feed page unsuccessfully queried if an error occurred.

The core data structure is the feed (or timeline); a set of feed items. A feed item is modeled such that any type of textual activity can be represented, i.e. posts, comments and stories. Obviously, stories play an important role in user feeds. Note, however, that stories often appear right next to the actual activity, especially for comments; therefore, the content will be duplicated without care. So as to not lose too much information when handling different types of feed entries, a few additional properties are needed to the standard Facebook set. In order to simplify the data model differences in the available post types are mostly ignored. Post types are links, photos, statuses, comments, videos, swfs (flash objects) and check-ins as well as the corresponding stories. After analyzing the properties of these entries, the following attributes were selected: the unique Facebook ID, timestamps representing when the entry was created and when it was last updated, the originator of the entry, optionally also the receiver of the entry and the comment and like count if present.

The originator and receiver are represented as separate resources, hence, only their unique IDs are stored here. The count of comments and likes are taken from the comments and likes properties of the Facebook format if present. A normal post has an attribute message which holds the text the user posted. A story, however, does not have a message, but rather a story property. The different sub-types of a post additionally have attributes containing the link, photo URL, etc. Each of these properties are mapped onto a single property. In order to distinguish between different types of feed items this property can be any of message, story or comment. The attribute then holds either story or comment for these two data types and the concrete post type for messages. A foreign key to the resource which this feed item belongs to, i.e. on which timeline it is posted. Last, to link comments to their respective post, a parent property is included, which is null for top-level posts.

5.4 Application of a Social Observatory: Political Sentiment in Germany

The initial use case of a Social Observatory analyzes political discourse and the expression of well-being in Germany. Politicians can serve as societal opinion makers and with the use of online social media, the potential for influence only grows. This study reviews 54,655 posts and 231,147 comments by 257,305 unique users at three granularity levels (all posts and comments per party; monthly posts and comments per party; individuals' posts and comments per party) in the year preceding and one month after the 2013 Federal elections. Users who only liked a politician's Facebook page (passive actors) are disregarded for lack of content. Macro trends are established, leading to discussions on the difference between politicians and constituents. The meso-analysis concentrates on discourse related to campaigning and expressions of communal cohesion, where the micro-level reveal individual well-being discourse patterns. Each granularity level of the Social Observatory reveals telling yet sometimes-contradictory indicators.

A convenience sample of the 620 members of the 17th German parliament (considering whether they have a publicly available Facebook account or not), found 190 politician with an open profile or page on Facebook, representing approximately 30% of Parliament. 187 had open pages, where data was fully publically available. Post refers to text pushed by politicians; comments refer to responses by constituents and politicians themselves. Table 5.1 illustrates some representative aspects of the dataset.

Table 5.1: Descriptive attributes of dataset, numbers are rounded for representation purposes

Party	Proportion of 17th German Bundestag	Proportion of Facebook dataset	Posts	Comments	Likes	Audience[29]
Grüne	11	11	6,586	41,744	194,528	38,665
CDU/CSU	38	40	20,006	68,667	493,891	119,212
FDP	15	11	4,835	26,703	118,215	21,046
Die Linke	12	13	8,886	26,471	178,816	24,986
SPD	23	25	14,342	67,562	501,483	80,300
Total	**100**	**100**	**54,655**	**231,147**	**1,486,933**	**257,305**

The synchronization of all active politicians in that group took 26:11 hours with no previously saved data, i.e. all data having been cleared before. The **UserDataGatherer** took 18:21 hours, which approximately refers to the time necessary for fetching the data, while transforming and

[29] Audience relates to the number of unique Facebook IDs that interacted with one or more politicians. Note: the total audience is not the sum for each party indicating that Facebook users interact with more than one party.

saving it to the database 7:50 hours. After 4:22 the Facebook query limit was reached for the first time. 31 minutes later operations could be resumed. In total the limit was crossed 13 times, on average after querying for 1:59. The average wait time until receiving new data was 24 minutes. Thus, 5:20 were spent waiting for the query limit to pass. The size of the Facebook resources varies greatly with the senior politicians like Angela Merkel or Sigmar Gabriel having tens of thousands of entries in their feed while less popular or newer members of parliament only have a few hundred posts and comments on their page.

The metrics calculations took 19:33 hours. One of the things early tests and subsequent improvements yielded was adding indices to all referenced fields in the data tables used for detecting shared resources. This alone yielded a speed improvement of around 50 per cent. In total the runtime did not decrease however, as more and more metrics were added to the set. At this point seven metric classes process a resource group and look at all possible 2-combinations. For the size of 190[30] resources this means 19,555 interaction points. Synchronization time is roughly linear with the number of resources, with the average time to fetch a resource ranging from five to eight minutes for the different pages. The metrics calculation, however, displays a clear non-linear relation ranging from twelve seconds per resource for the smallest and more than six minutes per resource for the largest group. This is due to the 2-combinations which have to be processed for a group, which scale non-linearly. Hence, it might be worthwhile to reduce the input to only include relevant profiles in order to increase the runtime and get closer to a real-time analysis. The Facebook data stored needed 798,784 KiB and the metrics tables used 90,132 KiB, about 3.5 MiB of data per resource.

Figure 5.3 visualizes interactions between politicians and their audience, capturing 85,679 bi-directional edges considering only text-based interactions, 345,704 considering only likes, and 385,936 when considering both. On average, politicians and their audience interacted 2.70 times, with a maximum of 1,503 interactions; 4.30 and 998 interactions respectively for likes, and 4.45 and 1,554 interactions considering both. Interactions between politicians are relatively low: there are 3,883 occurrences (0.23%) across all profiles. This suggests that Facebook is used mainly as a medium for promoting individual political agendas. Politicians posted on average 292 times. The average profile contains 29,301 words, from which 25% were six letters or more (a measure of linguistic variety).

[30] There were 190 politicians in the group, but three have unused profiles and are subsequently discarded.

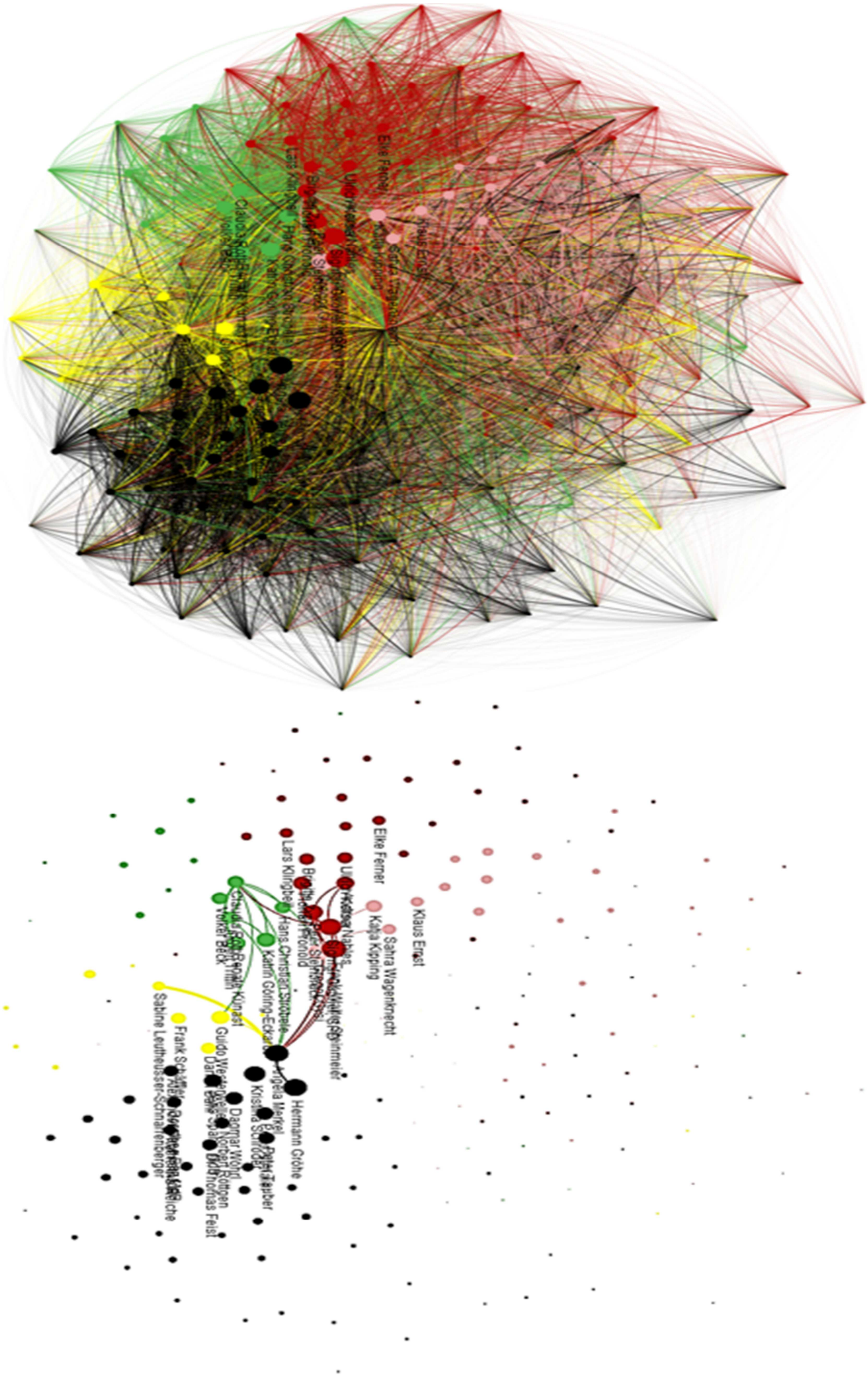


Figure 5.3: The extracted social interaction graph with all (a) and weightiest edges (b)

Figure 5.4 depicts the continuum of hourly posting behavior, with politicians posting in the morning and at lunchtime, and constituents responding in the afternoon. Politicians also tend to post on working days, whereas constituent volume shows no significant difference between weekdays and weekends (Figure 5.5).

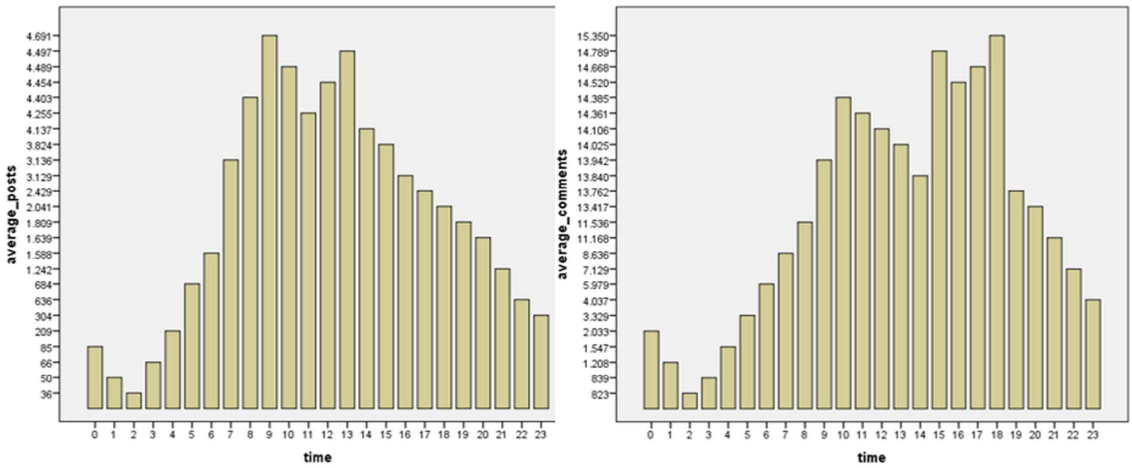


Figure 5.4: Distributions of hourly posting behaviors, posts and comments

The average post length was 40.8 words, differing from the findings of (Kramer 2010) who found that the average length of a Facebook post is nine words. This finding and its discrepancy compared to Kramer's results may have its origin in the particularity of this user sample. It is however a positive discrepancy, as the additional volume of text minimizes bias that could be incurred by low-volume (González-Bailón et al. 2014).

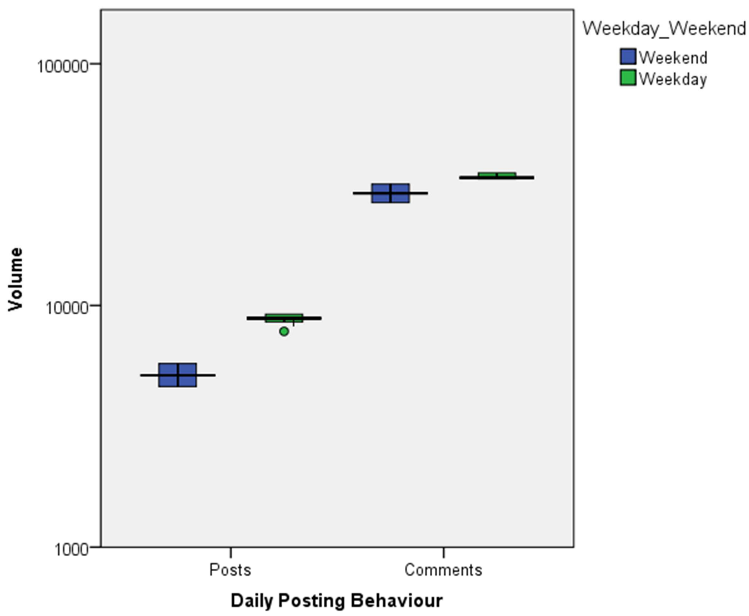


Figure 5.5: Weekday and weekend post and comment activity (logarithmic scale)

The monthly distribution of posts and comments depicted in Figure 5.6 show an increase in activity leading to the elections with two exceptions: a drop in December 2012, which was also observable in posts from 2009-2012, and a slight drop in July 2013 of posts by politicians, which is during the summer recess of the German Parliament. Posting activity significantly dropped in October 2013, directly after the elections. This drop is not reflected in the comments, nor is the recess drop in July. December is also a “slow” period for comments. Comments show spikes in November 2012 and March 2013, corresponding to interest in the various public scandals of the former President of Germany, Christian Wulff.

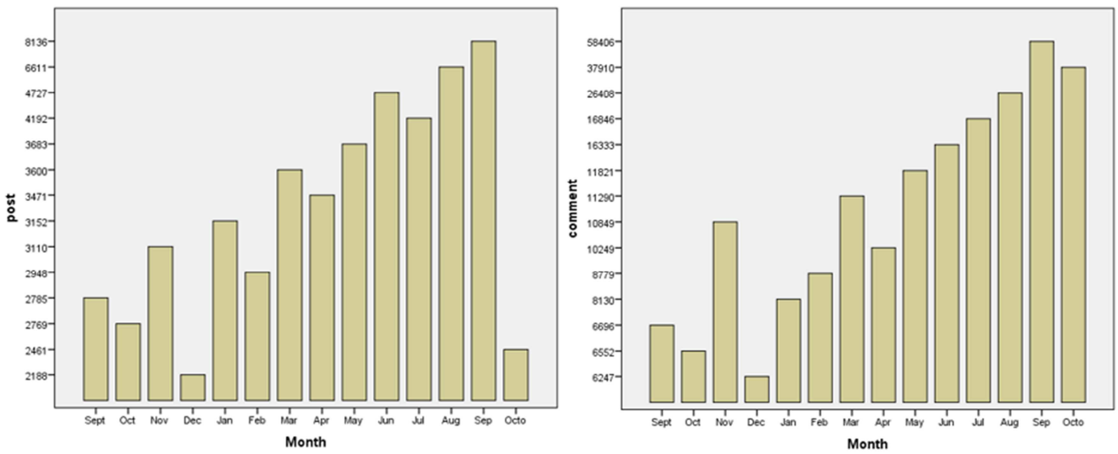


Figure 5.6: Total monthly posts and comments

The most commonly repeated post was “STOPPT die Massentötung in Rumänien! STOPPT die Tatenlosigkeit aller Verantwortlichen in der EU! JETZT!“ (*Stop the mass murders in Romania! Stop the inaction of EU stakeholders! Now!*), referring to Romanian ‘fur farming’ or domestication of animals for use in fur goods. 117 unique users, 234 times in total, repeated this single post.

5.5 Evaluating a Social Network at Multiple Resolutions

5.5.1 Macro-level Assessment

In order to assess the (dis)similarities of language between the parties and their constituents, the study employs the nearest neighbor method and with simple Euclidean distance classifies the similarity of the samples between parties, constituents, and parties and their constituents. The attributes of the feature vector are the individual LIWC scores per sub-group. This allows a more textured view of German political discourse on Facebook. For two instances in a general *n*-dimensional space:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{5.1}$$

where the distance *d* between instances *x* and *y* results from the square root of the sum of the squared differences between the values for the cases, over all dimensions. Similar cases are near to each other whereas cases with low similarity are far(ther) apart. The upper limit of distance is dependent on the size of the hyperplane. Therefore, the distance between a given pair can be used as a measure of their (dis)similarity. 45 unique permutations of posts and comments from the parties and their audiences exist for 64 LIWC variables, creating a 64-dimensional space. Each instance (n=10) is one centroid representing a party's posts or comments. The centroid is a hyper plane calculated based on the centroids of each instances' 64 LIWC variables. Mimicking the method of (Pang and Lee 2005) supervised learning from the training set is reported (Table 5.2).

Two issues necessary to consider when dealing with hyperdimensionality are the "curse of dimensionality" and "hubness" (Radovanovic, Nanopoulos, and Ivanovic 2010). The distance between comments and posts is small (considering that this is a 64-dimensional plane), with an absolute range from 2.017 (Linke comments and SPD comments), to 10.523 (Grüne posts and SPD comments) (Table 5.2). As the space is small but not equal, high dimensionality was not found to unexpectedly compress the data. As there are no "popular" hubs, it can also be rejected that hubness is driving these results.

All comments are closest to other comments and all posts are closest to other posts. Comments are more similar to each other than posts. Whereas the absolute distance between comments is [2.017 – 4.665], the range between posts is [4.140 – 6.645]. Distance is revealing: e.g., politicians from the CDU/CSU and SPD are expected to be dissimilar but rather are one another's nearest neighbors, while governing block members largely do not occupy the same space. Only the SPD and Grüne have party and constituent closeness at k=5, but this is not the case for the CDU/CSU, FDP, or Linke. In no case is a party-constituent pairing closer than k=5. The governing blocks' language patterns are largely intransitive.

Table 5.2: Nearest neighbors where k= 5, politicians and constituents

	k=1	**k=2**	**k=3**	**k=4**	**k=5**
CDU/CSU comments	Grüne (4.082)	SPD (4.209)	FDP (4.303)	Linke (4.655)	Grüne_p (10.487)
Linke comments	SPD (2.017)	Grüne (3.170)	FDP (3.413)	CDU/CSU (4.665)	FDP_p (10.156)
FDP comments	Grüne (3.050)	Linke (3.413)	SPD (3.461)	CDU/CSU (4.303)	Grüne_p (10.156)
Grüne comments	FDP (3.050)	Linke (3.170)	SPD (3.210)	CDU/CSU (4.082)	Grüne_p (9.872)
SPD comments	Linke (2.017)	Grüne (3.210)	FDP (3.461)	CDU/CSU (4.209)	FDP_p (9.982)
CDU/CSU posts	SPD (4.140)	Linke (5.201)	FDP (5.507)	Grüne (6.041)	SPD_c (10.523)
Linke posts	SPD (4.386)	FDP (4.645)	CDU/CSU (5.201)	Grüne (6.089)	SPD_c (10.523)
FDP posts	Linke (6.645)	SPD (4.730)	CDU/CSU (5.507)	Grüne (5.870)	SPD_c (9.982)
Grüne posts	FDP (5.870)	SPD (5.898)	CDU/CSU (6.041)	Linke (6.089)	Grüne_c (9.872)
SPD posts	CDU/CSU (4.140)	Linke (4.386)	FDP (4.730)	Grüne (5.898)	SPD_c (10.184)

While the results above indicate that the feed patterns found in political discourse largely occupy the same space, a paired sample t-test finds that overall the five parties do have differences in feed patterns as represented by their respective LIWC categorizations. Again, 64 LIWC sentiment categories are assessed for 45 unique party-constituent permutations. There are statistically significant differences in 35 political party and audience pairings out of the possible 45. All results are available for review in Appendix III.

While some results are not unanticipated, other pairings are unusual. There is no significant difference between the posts or comments of the two center-right parties CDU/CSU and former coalition partners FDP ($t(63)$ -1.788, $p < .05$), or between the leftist parties SDP and Linke ($t(63) = -.290$, $p < .05$). In addition, no significant differences between the posts and comments of either the center-right CDU/CSU or FDP and the socialist Linke party ($t(63) = -.893$, $p < .05$); ($t(63) = -.867$, $p < .05$) are found. Interestingly, the only non-significant difference of the Grüne was between that of the posts of the CDU ($t(63) = .799$, $p < .05$). All other pairings with the Grüne were significantly different. It must be noted that all post-comment combination have significant differences, which is supported by the results of the nearest neighbor test.

These differences between relationships as found in the nearest neighbors and t-tests are interesting, as it suggests that politicians and their audiences on Facebook could be concentrating on different points, or are giving importance to different topics across their general discussions. Alternatively, this finding supports the assumption that there is a diversity of political conversation amongst Facebook users. As the parties are platform based, this is a positive finding. The results defy the thesis of linguistic accommodation of (Niederhoffer and Pennebaker 2002); a reason for the lack of coalescence here be could that conversation partners change too rapidly to adapt to one another. It is worth noting that the overall corpus follows the pattern of polite discussion put forth in (Brown and Levinson 2013; Pennebaker, Mehl, and Niederhoffer 2003).

With regards to expressions of positive and negative emotions (well-being) rather than the entire spectrum of sentiment, the results are contradictory to those above. In order to benchmark the politicians' posts against party norms, LIWC assessments of the most recently published party manifesto are included, represented by the party name only. Figure 5.7 shows the relationships graph following the calculation of dissimilarity of Equation 5.1, where edge thickness as well as centrality represents similarity in latent well-being expressions. Notable is that all manifestos are rather disconnected from their parties posts and comments, with the notable exception of the Grüne, whose comments and manifesto share similar dimensionality. The CDU/CSU, SPD, and FDP manifestos express well-being similarly.

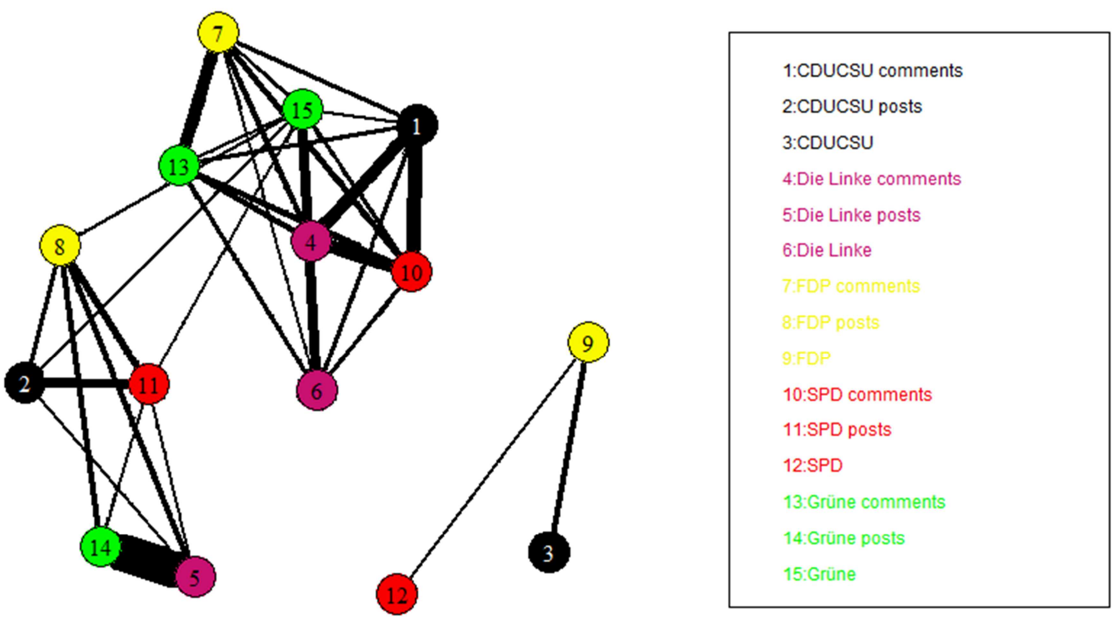


Figure 5.7: Expressed well-being relationship matrix, estimated via Euclidean distance

There is a notable cluster of posts on the left side of the graph; politicians are expressing well-being similarly across their posts. The strength of the similarly of the Grüne and the Linke could be explained in that they are the two 'minor' parties in the opposition, and thus are reaching out particularly emotionally to their constituents. Especially dissimilar in expressed well-being are each the CDU/CSU and FDP, as seen by their lack of intra-party connectivity and relatively high distances across manifestos, posts, and comments.

5.5.2 Meso-level Assessment

Social metrics derived from differences in LIWC categories reveal the patterns of discourse (C. Chung and Pennebaker 2011; Pennebaker, Mehl, and Niederhoffer 2003). Obvious in this dataset is a distinct propensity to discuss in present tense, which can suggest either that politicians on Facebook are not in fact ‘campaigning’ in the traditional sense, but are rather discussing daily life with their constituents; or, that verbal immediacy (familiarity) is in place (Mehl and Pennebaker 2003). With respect to the analysis of communal well-being, either assessment can be seen as a sign of community building, or the fostering of online positive relationships and communal belongingness (as defined in the terms of Human Flourishing).

The findings reported in (Tausczik and Pennebaker 2010) of a political discourse study by Gunch and colleagues (2000) states that this could also be related to positive campaigning rather than ‘dirty’ campaigning. Manifestos have 3.19 times more references the present than the past and 3.05 times more references the present than the future, with the exception of the Grüne manifesto that has an inverse present-future relationship. Posts are slightly more balanced with present/past references having a 1.57 difference and present/future discrepancies at 2.73. Comments are the most present-focused, with audiences referring to the present 3.23 times more than the past and 4.46 times more than the future. Considering the population, this is an unexpected finding. Whereas it may not be unusual for politicians and political discourse to focus on the present rather than the past, the absence of future references, especially in the face of national elections, is unanticipated (Figure 5.8).

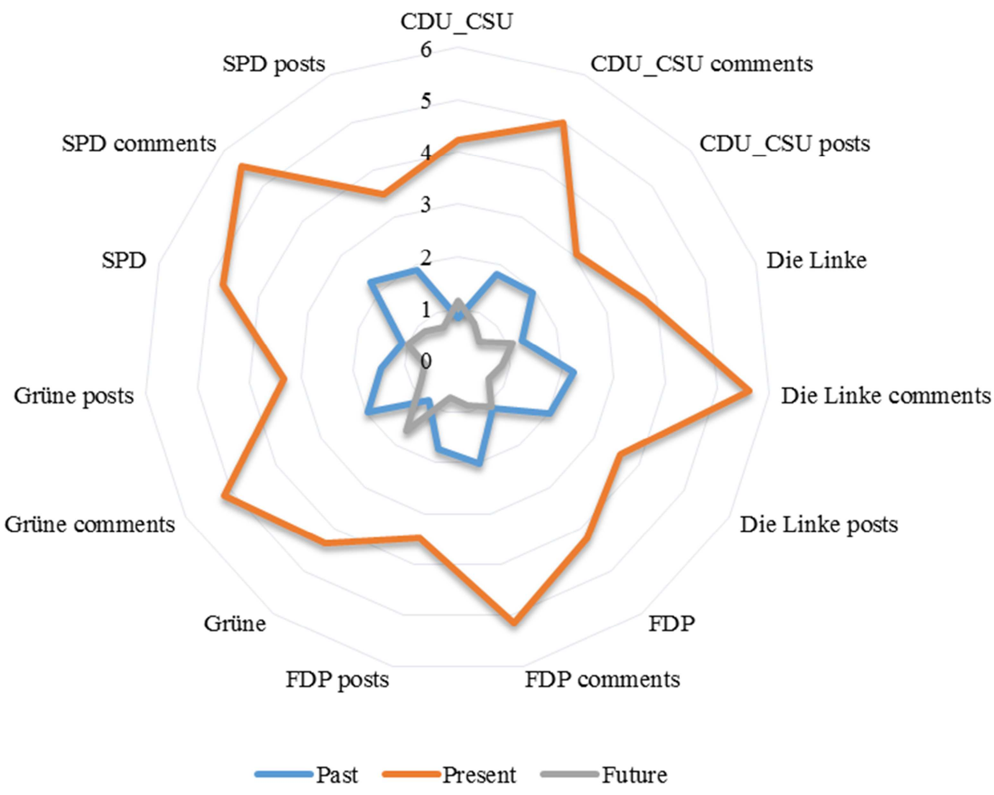


Figure 5.8: Language tense patterns of party manifestos, posts, and comments

Political discourse does seem to be communal discourse as displayed by the manifestos and Facebook activity. Social references rank well above references to the self; first person plural and the second person "you" come before first person singular (Figure 5.9). Considering a visual analysis of the data, there is no cause to believe that the politicians or constituents are using the "Royal We," in which "we" is used to imply cohesion but indicates commands (Tausczik and Pennebaker 2010). This tendency towards communal discourse can be seen as an indication of communal belongingness (a positive well-being indicator) as defined in Chapter 3.2.3.

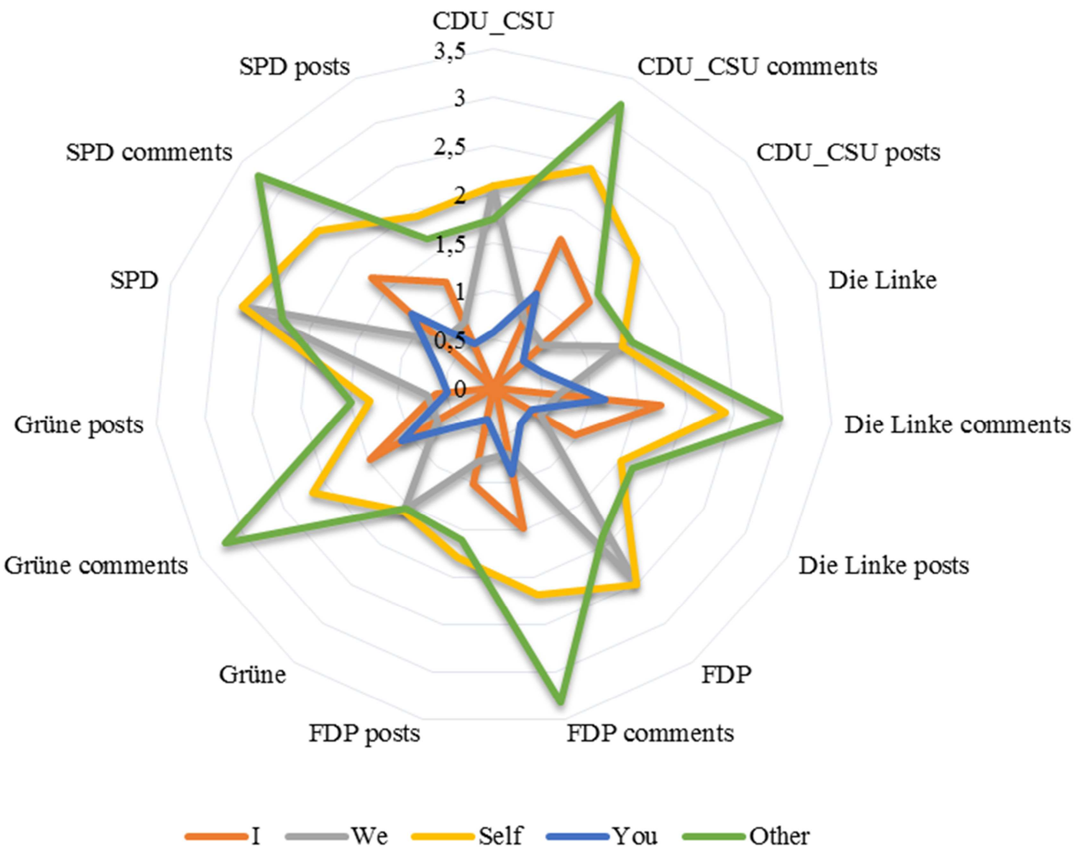


Figure 5.9: Social references in party manifestos, posts, and comments

This work finds no significant correlation between positivity, negativity, use of first or third person, and tense and thereby cannot replicate (Gunsch et al. 2000), who state that first person references are related to positive campaigns and third person campaigns are related to negative campaigning. Also rejected is that the social accepts of feed reflects an "Us-Them" mentality, when taking the relative frequency of inclusivity and exclusivity into consideration (Figure 5.10). Especially manifestos and posts orient towards inclusive discourse. Comments, whilst having spikes of exclusionary sentiment, are also overarchingly inclusive. This again supports the concept of communal belongingness as an indicator of positive well-being.

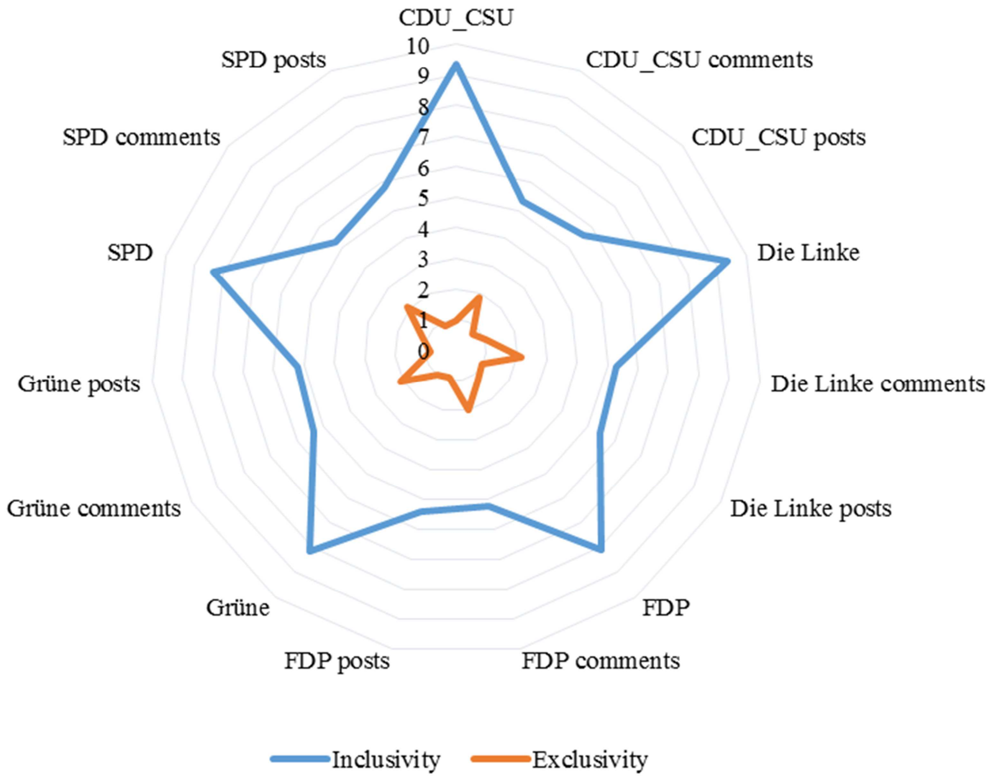
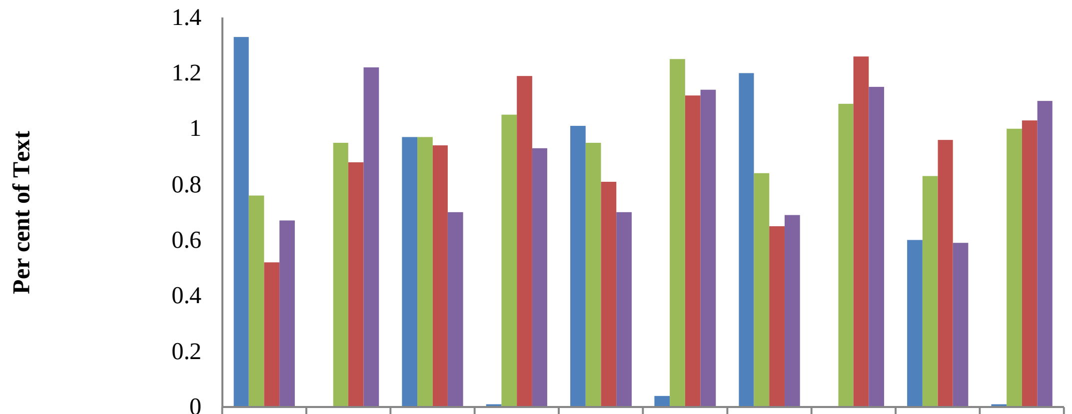


Figure 5.10: Inclusion and Exclusion references in manifesto, posts, and comments

Additional interesting patterns in these samples are observable. Negative emotions, anger and money discussions are positively related (rs(331) = .137, p<.0005; rs(331) = .184, p<.0005), reflecting on-going public sentiments at bailouts to neighboring countries. Optimism, positive emotions and achievement also have a positive relationship (rs(331) = .362, p<.0005; rs(331) = .306, p<.0005).

A tempting item to evaluate is the presence of deception, defined by (Newman et al. 2003; Pennebaker, Mehl, and Niederhoffer 2003) as usage patterns of higher negative emotion, more motion words, fewer exclusion words, and less first-person singular. Western cultural stereotypes are replete with the image of political misrepresentation – does this hold up to empirical analysis? The macro analysis finds that no single subgroup has a profile indicative of deception (Figure 5.11), indicating that as a whole, parties are posting quite honestly about their activities. This is in line with the previous finding, as if politicians are discussing their and their constituents activities, there is little incentive to lie. It must be noted here that individuals could have quite different profiles; at the aggregate though, it is not justifiable to continue zooming into individual profiles.

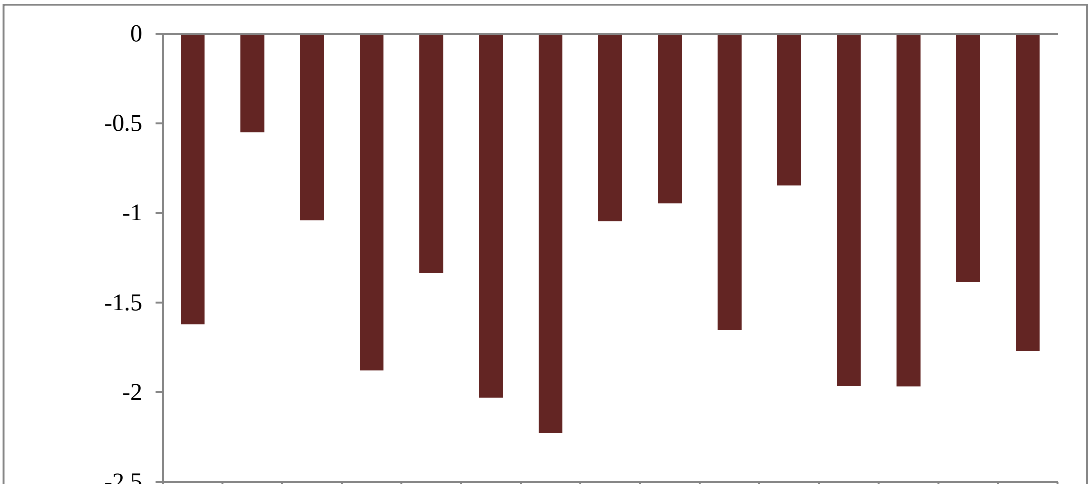


	CDUCSU_posts	CDUCSU	Die_Linke_posts	Die_Linke	FDP_posts	FDP	SPD_posts	SPD	Grüne_posts	Grüne
I	1.33	0	0.97	0.01	1.01	0.04	1.2	0	0.6	0.01
Exclusion	0.76	0.95	0.97	1.05	0.95	1.25	0.84	1.09	0.83	1
Negative emotion	0.52	0.88	0.94	1.19	0.81	1.12	0.65	1.26	0.96	1.03
Motion	0.67	1.22	0.7	0.93	0.7	1.14	0.69	1.15	0.59	1.1

Figure 5.11: Percentage of words in a deceptive profile, per party across manifesto, posts and comments

5.5.3 Micro-level Assessment

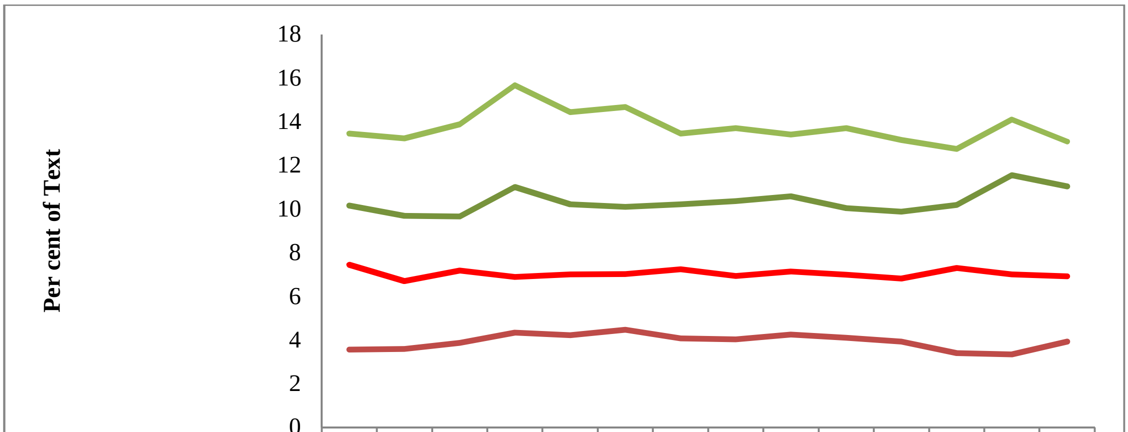
While warning scholars to proceed with caution, (Pennebaker, Mehl, and Niederhoffer 2003) identified positive and negative sentiment analysis as an area of future research in their 2003 Annual Review of Psychology article. As expected, emotion words in the corpus are relatively low, accounting for 0.11 - 4.2 per cent of all posts or comments. As the experience of positive and negative emotions is formative to well-being (Diener et al. 1985; Huppert and So 2009), positive and negative sentiment are still evaluated as a singular item of focus. One common method to identify the ‘baseline’ of written positive and negative emotion is to subtract negative sentiments from positive sentiments (Kahneman and Krueger 2006; Kahneman et al. 2004b). When applying the LIWC dictionary, this requires grouping the variables Positive Emotion, Positive Feelings, and Optimism as well as the variables Negative Emotion, Anxiety, Anger, and Sadness. Subtracting the negative emotional categories from the positive results in the variable ‘Net Affect.’ While ‘Net Affect’ is highly correlated with the existing LIWC category Affect ($r_s(275) = .763$, $p < .0005$), they reflect different word usages according to the LIWC dictionary. Net Affect is therefore a more diverse measurement of positive and negative emotion. Interestingly, the Net Affect of political discourse on Facebook is negative (Figure 5.12). Considering that this study takes place in advance of an election year, this display of negative sentiment is rather unexpected. As seen in the coming figures, this indicator is too highly aggregated. The measure of simple positive and negative emotion has much more telling and specific features.



	CDU CSU comments	CDU CSU posts	CDU CSU	Die Linke comments	Die Linke posts	Die Linke	FDP comments	FDP posts	FDP	SPD comments	SPD posts	SPD	Grüne comments	Grüne posts	Grüne
Net Affect	-1.62	-0.55	-1.04	-1.88	-1.33	-2.03	-2.23	-1.05	-0.95	-1.65	-0.85	-1.96	-1.97	-1.39	-1.77

Figure 5.12: Net Affect of German Political Discourse on Facebook

Summing all posts and comments, then analyzing for monthly changes results in the graph depicted in Figure 5.13. The rise in positive sentiment within the last month of 2012 is due to increased use of holiday wishes analogous to the finding of (Dodds et al. 2011; Kramer 2010). An additional bump in positive sentiment for both posts and comments is visible coinciding with the lead up to the federal elections, along with a minor drop in negatively intoned posts.



	2012_09	2012_10	2012_11	2012_12	2013_01	2013_02	2013_03	2013_04	2013_05	2013_06	2013_07	2013_08	2013_09	2013_10
Sum Pos Posts	10.16	9.68	9.65	11	10.21	10.1	10.21	10.36	10.58	10.03	9.87	10.18	11.54	11.03
Sum Neg Posts	3.56	3.58	3.87	4.34	4.22	4.47	4.07	4.03	4.24	4.1	3.92	3.39	3.34	3.93
Sum Pos Comments	13.45	13.23	13.88	15.66	14.44	14.67	13.45	13.7	13.41	13.7	13.16	12.75	14.09	13.08
Sum Neg Comments	7.44	6.7	7.18	6.88	7	7.02	7.24	6.93	7.13	6.99	6.81	7.29	7	6.91

Figure 5.13: Average positive and negative sentiment per month, posts and comments

As seen in Figures 5.14a-d, positive and negative sentiment at the party-level and user-level is even more distinctive. The greater use of words bearing positive sentiment compared to words bearing negative sentiment is noticeable, especially in light of 60% more words within the LIWC dictionary being associated with negative sentiment (Wolf et al. 2008; Pennebaker et al. 2007). Overall, manifestos have nearly double the occurrence of positive emotion words as compared to posts and comments, and are more negatively intoned than posts in all cases. Positive sentiment within the posts and comments often concern congratulations on birthdays, campaigning activities, and self-promotion. This suggests that the message that the parties would like to display is not necessarily being followed in day-to-day interactions of politicians and their constituencies.

At this granularity level, there are almost no differences in the means of negative emotion usage, with posts tending to contain slightly less negative emotion words as compared to party manifestos and comments. This is also reflected in Figure 5.12, where posts are consistently the least negative of all observations, as well as Figure 5.7, where posts are the most tightly clustered group. A visual inspection found that posts and comments high in negative sentiment typically detail concerns about child abuse, night flight operations, as well as the situations in the Middle East and the financial situations with Greece. This is supported by the correlations between negative emotions and references to money. While criticism of opposing parties is present, the low negativity levels suggest that ‘dirty’ campaigning on Facebook is kept to a minimum. As the comments are both more positive and more negative this suggests that there is a minimum of self-promoting behavior, or narcissism, amongst politicians (Davenport et al. 2014).

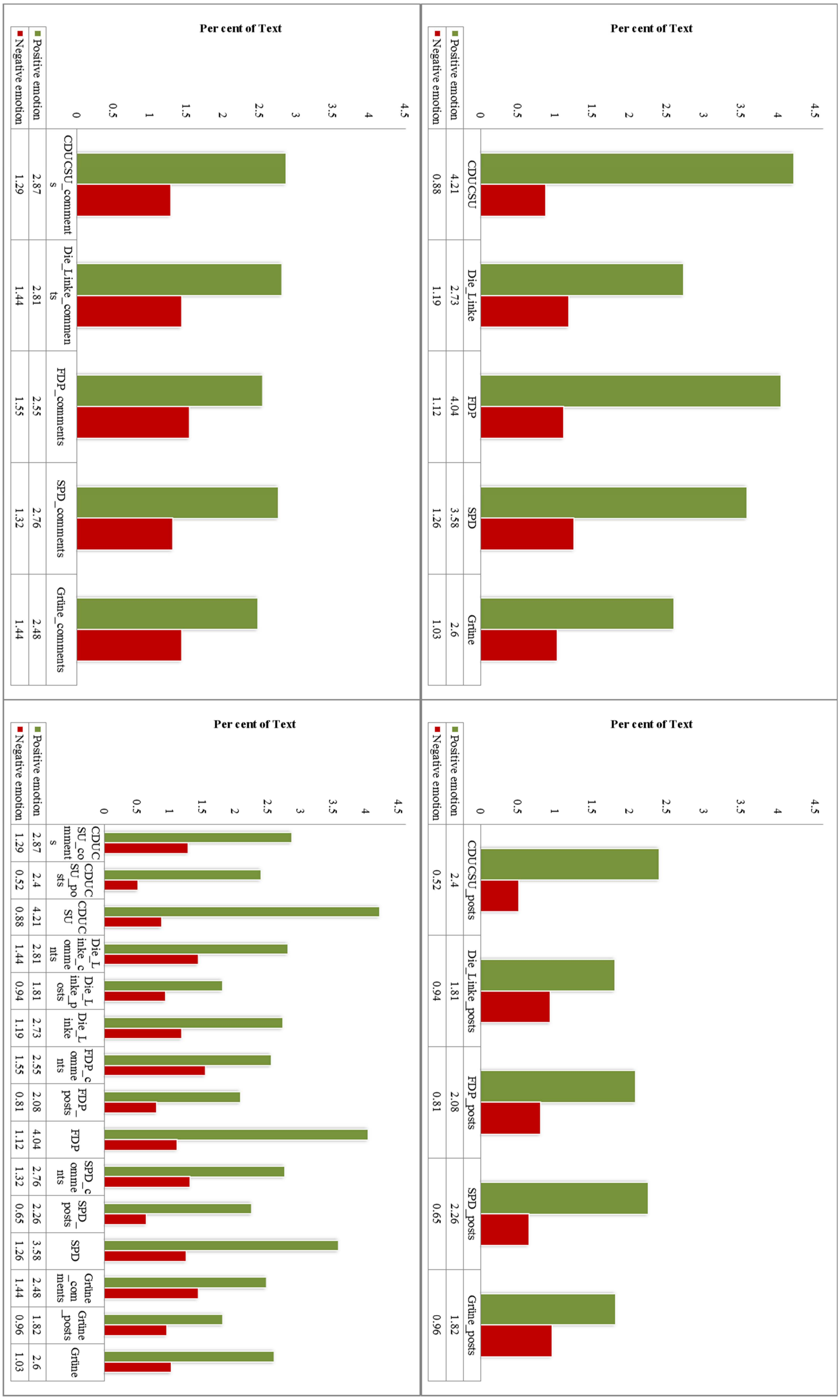


Figure 5.14: Sentiment by (a) Manifesto, (b) Politicians, (c) Constituents, and (d) Overview of all

From this micro-disaggregation, it becomes possible to see which politician has the most negative and positive dialogue per party (Table 5.3). An interesting feature here is found in positive and negative comments. While there are not significant differences at the party aggregate level, the top five positive commentaries are directed at CDU/CSU politicians, and four of five most negative commentaries are directed at the Linke. Another notable feature is that while posts from Peer Steinbrück, the SPD contender for Chancellor, are amongst the most positive, Chancellor Angela Merkel appears neither in the most positive nor negative posts and comments. That Marieluise Beck has the most negative posts of the entire dataset is not unexpected as her platform includes criticisms of environmental policy and human rights abuses across Europe along with her known status as a vocal critic of the Russian leader Vladmir Putin. Ms. Beck's Facebook discourse gives context to the stance of well-being scholars that the experience of negative emotions is not a bad thing, and in fact is necessary for the development of well-being (Ryan and Deci 2001; Diener et al. 1999)

Table 5.3: Most positive and negative posts and commentator groups by relative per cent

Name of Politician	Party	Positive	Negative	Party	Name of Politician
Günter Glose posts	SPD	6.17	1.66	Grüne	Marieluise Beck posts
Ingo Wellenreuther posts	CDU/CSU	3.62	1.65	CDU/CSU	Ernst-Reinhardt Beck posts
Hens Peter Friedrich posts	CDU/CSU	3.61	1.54	Linke	Ulla Jelpke posts
Peer Steinbrück posts	SPD	3.59	1.54	Grüne	Omid Nouripour posts
Franke Edgar posts	SPD	3.53	1.40	CDU/CSU	Guido Westerwelle posts
Gero Storjohann comments	CDU/CSU	9.9	3.85	Linke	Andrej Hunko comments
Albert Rupprecht comments	CDU/CSU	8.78	2.75	Linke	Karin Binder comments
Peter Wichtel comments	CDU/CSU	8.64	2.04	SPD	Sascha Raabe comments
Ewa Klamt comments	CDU/CSU	8.47	1.97	Linke	Dorothée Menzner comments
Sabine Weiss comments	CDU/CSU	8.31	1.88	Linke	Richard Pitterle comments

Similarly, at this granularity it is possible to view the politicians and constituents indicating the highest tendencies towards inclusion and exclusion (Table 5.4). This seems to have little relationship with election results, as only three politicians did not re-join the 18th German Federal parliament, although Ms. Höll (exclusionary commentators – 4.62%) did lose her position in parliament.

Table 5.4: Most inclusive and exclusive posts and commentator groups by relative per cent

Name of Politician	Party	Inclusion	Exclusion	Party	Name of Politician
Sascha Raabe comments	SPD	8.57	4.62	Linke	Barbara Höll comments
Claudia Roth posts	Grüne	7.63	3.2	Linke	Dorothée Menzner comments
Franke Edgar posts	SPD	7.62	2.86	SPD	Anette Kramme comments
Diether Dehm posts	Linke	7.52	2.86	Linke	Diether Delm comments
Günter Glose posts	SPD	7.41	2,75	Linke	Karin Binder comments
Ernst-Reinhard Beck comments	CDU/CSU	7.32	2.63	CDU	Hans-Joachim Fuchtel comments
Sibylle Pfeiffer posts	CDU/CSU	7.24	2.55	SPD	Petra Ernstberger comments
Aydan Oezoguz posts	SPD	7.23	2.54	FDP	Daniel Volk comments
Frank Walter Steinmeier posts	SPD	7.12	2.53	Grüne	Friedrich Ostendorff posts
Rainer Arnold posts	SPD	7.01	2.53	CDU/CSU	Gunther Krichbaum comments

A further look at social discourse between individual politicians to their constituents bears final interesting features. At the politician level, this work found no significant differences in discourse patterns based on gender, nor are there gender differences found in constituents' responses to politicians. Posts tend to be statements and comments tend to ask questions, which could be indicative of the finding that higher status people ask less questions (Tausczik and Pennebaker 2010). Anecdotally, Chancellor Merkel's posts did not contain a single question mark for the 13 months of this analysis.

5.6 Discussion

German political discourse is a rich, dense network. German political discourse occupies a close space, though distinct characteristics and relationships appear when viewed at the correct resolution. A tempting assessment is that the use of Facebook data for analysis between politicians is unnecessary, as it is signaling cohesion between their platforms. However, one overarching fact of this study is that posts and comments are oftentimes intransitive, indicating that politicians and constituents are more often than not talking past one another. While the two largest parties (CDU/CSU and SPD) tend to use online feeds in similar ways, the three smaller parties have attributes onto themselves. Where the Grüne is the least similar and most future-oriented party, the Linke has the highest concentration of negative commentators.

Distinct in its nondescriptness, the FDP showed no discrete patterns. This lack of distinctiveness is quoted as a major reason why the FDP did not meet the minimum criteria of to be re-elected into the 18th Parliament.[31]

Positive and negative sentiment are interesting indicators in terms of communal mood, but show only limited potential as public opinion gauges. This is due to the missing component of personality – without an estimate of aspects like extraversion and neuroticism as established with **RQ 1.1**, a baseline of well-being is difficult to establish. This lack of benchmark is also closely related to a limitation of this chapter; the need to cross-verify the data with study participants. Much more revealing is the sentiment analysis in its entirety (**RQ 2.3**). Discourse on Facebook is polite yet hierarchical, and outside of gendered discourse. Aspects of communal belongingness and familiarity are found. Facebook offers an open, deliberative and participatory civil society forum for exchange. Active campaigning is kept to a minimum, in favor of continuous updates of how the politician is serving the community. However, where politicians seek to be as inclusive as possible, constituents are careful to make distinctions in their viewpoints. Interesting to investigate in the future is to what extent this impacts communal belongingness. While differences are fine at the coarsest level of analysis, patterns can be detected. Sentiment analysis at a user-level is promising, as aggregating sentiment levels of users to a higher party average or overall average leads to an averaging value without distinct significance, causing a blurred view. Accordingly, it is striking that when observing at different levels, i.e. all, a party, or an individual, subtleties otherwise lost in the aggregation method are uncovered. Individual sentiment scoring is an especially poignant method for a TSR application. This was illustrated in the lack of gendered discourse and gender-directed responses in the face of a growing body of literature stating that Internet anonymity can increase sexist remarks.[32]

This analysis of political sentiment mining indicates that modern assessments of public opinion are largely improperly scaled. It cannot be understated that standard national indicators in use today rely on the aggregated view and not that of the individual or (sub)group. This supports the argumentation of Chapters 1 and 3. It also partially fulfils the requirements set up in **RQ 2.3**. By correlating public sentiment with other data like location, socio-economic data, age, political party or others, researchers and decision makers can begin to identify and categorize the impact of political actions. The value of the Social Observatory approach is also that it is use case independent: approaches outside of well-being like crime tracking, event prediction, and institutional monitoring are easily within scope.

---

[31] http://wahl.tagesschau.de/wahlen/2013-09-22-BT-DE/analyse-wanderung.shtml#11_Wanderung_UNION (infographic in German) Last Accessed: 11 November 2013.

[32] http://www.theatlantic.com/technology/archive/2014/10/the-unsafety-net-how-social-media-turned-against-women/381261/. Last Accessed: 20 October 2014.

In policymaking, public conversation and governments sometimes face a chasm. The Social Observatory monitors both the public mood on policy implementations, and possible negative backswings. It also has the ability to cluster public text in a way which both highlights similarities and differences between parties and audiences. Emphasizing current topics of conversation is also not to be undervalued in the era of the 24-hour news cycle, where the flashiest information is oftentimes the most frequently shown, even if they are not the topics which are in discussion around the dinner table. In this way, Social Observatories contribute an expansion of the methodology for empirical TSR applications (**RQ 1.2**).

5.7 Limitations and Conclusion

Whilst the results of this case study are encouraging, the methods are not without fault. Within the quality control of selected users posts with incorrectly labelled sentiment scores were identified. Those deviations can have different reasons. A misinterpretation by the word/word stem approach is most likely, as these methods are notoriously hard to apply to cases of ironicism and sarcasm (Tsur and Rappoport 2010). The post filtering approach can be revisited: this exploratory study includes only status updates without photos, videos or links. Some politician profiles heavily use media content (e.g., Angela Merkel), and are consequently largely omitted from the analysis. Another issue is that politicians have PR teams that often post on their behalf. As such, the feature extraction and filtering methods should be extended to enable differentiated authors. This would require a nearly post-by-post analysis of latent sentiment patterns which is nearly impossible on a dataset of this size due to the tool in use. The text analytics functionality currently provided by LIWC is limited; making it a tool invocable from the command line for the future iterations of the Social Observatory workflow would be worthwhile.

The continuing integration of the offline and digital self creates new requirements for social researchers and stakeholders. As mentioned in the preceding section, whilst the Social Observatory is a useful method for the extraction of data and supporting of analyses, the current iteration is missing a feedback loop to study participants. This loop would enable the cross-verification of aspects like belongingness or well-being. Additional data like personality could be attained with such a loop; also verifiable would be if the discourse participants are employing alternative personas to embody an online idealized self (Hilsen and Helvik 2012; Buckels, Trapnell, and Paulhus 2014). The approach's current iteration does not allow for such secondary analyses, and as such requires further research (**RQ 2.4**). This is in fact important for a proper meso- and micro-level analysis, and should be considered in future iterations of the Social Observatory as well as in future TSR applications.

More and more, interactions and reactions to institutions happen online. Missing is a generalizable, open-source tool for accessing and analyzing these phenomena. This chapter presents the vision and architecture of a Social Observatory: a low latency method for the observation and measurement of social indicators within an online community. To explore the usefulness and possibilities of a Social Observatory for policy and decision makers, a Facebook adapter was implemented, focusing the Observatory on 187 German federal politicians and 257,305 lay constituents, as proxies to public opinion. User interaction is observable and by leveraging the LIWC text analysis toolkit, different facets of communication processes are identified and significant differences in sentiment between the politicians and their followers are observed.

The implications of this work are threefold; firstly, a framework to automatically extract public data troves (even from Facebook profiles) for use in studies related to online communities is created. Secondly, that with a few generalizable tools quite complex interdisciplinary research processes can be undertaken. Finally, using only a small number of points of reference, i.e. the 187 politicians, the approach can discover and analyses the actions of an entire (sub)community (**RQ 3**). By employing similar techniques and extending the analysis stages, undertaking the same study on any online social community is enabled, shedding light on specific social dynamics, and identifying key or influential actors unobtrusively. This ability is of key strategic use for public figures that wish to assess for example their public standing, or the reactions to specific actions.

Chapter VI Detecting Self-Representation and Well-being on Facebook

"When an individual appears before others his actions will influence the definition of the situation which they come to have. Sometimes the individual will act in a thoroughly calculating manner, expressing himself in a given way solely in order to give the kind of impression to others that is likely to evoke from them a specific response he is concerned to obtain."

The Presentation of Self In Everyday Life, (Goffman 1959)

It is indisputable that social media and the Internet reshaped information disbursement and processing. This leads to specific challenges in adapting to the management of communication. As a generalization, social media users can be split into two groups: users who search for information, and users who produce and/or form information (Auer 2011; Kushin and Yamamoto 2010). Especially important for researchers and practitioners is observing and managing the effects of information creators on information recipients (Auer 2011). Poorly created informational content can contribute to what is known as the 'spiral of silence' in public opinion, both on and offline (Hampton et al. 2014; Noelle-Neumann 1974). This need is more pressing in the face of recent findings from Pew Research, that 30% of Americans primarily receive their news from Facebook, 10% from YouTube, and 8% from Twitter (Hampton et al. 2014). Especially considering that oftentimes users actively search for opinions mirroring their own, the veracity of crowd-disbursed information is of upmost importance.

This veracity is a reason online social data raises challenges for researchers aiming to unobtrusively apply publically accessible online data to generalizable social models. As seen in Chapter 5, the trove of potential data is vast, but the ability of researchers to verify its veracity is low. Across platforms like Facebook, LinkedIn, Twitter, and blogging services, users (sub)consciously represent themselves in a way which is appropriate for their intended audience (Qiu et al. 2012; Zhao, Grasmuck, and Martin 2008). However, researchers have not yet adequately addressed controlling for self-representation in online social networks, or the propensity to display socially responding characteristics or censorship of oneself in online fora, (Das and Kramer 2013; Zhao, Grasmuck, and Martin 2008). As such, researchers on these platforms risk working with 'gamified', or socially responding personas that go beyond efforts to contain CMB (Linville 1985; Podsakoff et al. 2003; Ruths and Pfeffer 2014; González-

Bailón et al. 2014). What has not been approached in a systematic way is the verification of such data on offline and actual personality (this chapter uses the same definition of personality as in the preceding chapters). This leaves the open question of alignment of unobtrusively gathered data and online self-reported data. This chapter focuses on the alignment of survey methods with unobtrusive methods of gathering data from online social media in support of accurate assessments of the micro-level of the TSR framework.

The chapter hypothesizes that self-representation can be identified, and thus eventually be controlled for in broad social models (Section 6.1). This enables the social research to obtain online social media data and pre-process it accordingly for use in TSR models. For this study, the popular crowdwork platform Amazon Mechanical Turk (AMT) was employed. Survey responses and Facebook Timeline data from 509 workers (Section 6.2-6.3) were recorded. Sections 6.4-6.5 discuss and summarize the contribution, limitations, and points out areas for future work. This chapter is built upon and extends the working paper (Hall and Caton 2014), presented at the Internet, Policy, and Politics conference held at the Oxford Internet Institute.

6.4 Conceptual Background

Self-representation has been discussed in several works for online and offline fora. These studies discuss that one's tendency to truthfully disclose personal information emanates from an associated intrinsic value (Ellison, Heino, and Gibbs 2006; Lawson and Leck 2006; Mehra, Kilduff, and Brass 2001). Specifically personality and expression of well-being are interesting to assess for signs of self-representation due to their known relationships in on- and offline fora. While many methods including surveys, interviews, and (n)ethnographic research can identify self-representation from the first person perspective, text analytics is a promising research design for the unobtrusive identification and mitigation of self-representation bias in data at a lower overall cost.

6.1.1 Self-representation and Online Social Networks

Self-representation is distinct from the concept of identity contingencies (Purdie-Vaughns et al. 2008), where self-representation is the presentation of idealized self and identity contingencies is the presentation of a social identity marker (e.g., being a computer scientist, being from the United States). In real life direct communication is often the social norm (Hoever 2010) whereas in social networks communication is more indirect. Users present themselves online by means of likes, text, music, video and pictures. Status updates, uploading pictures or inserting information in the "About Me" section is not directed to anyone specifically. Although one approximately knows who may be reached, it is not known who will respond. As Facebook is not anonymous (in opposition to Twitter) the freedom of identity construction is significantly restricted. Most people use Facebook to stay in touch with people met offline, so they cannot completely detach their true identity (Zhao, Grasmuck, and Martin

2008). Thereby users try to present a socially aspired self-image to be 'popular' (Utz, Tanis, and Vermeulen 2012). In (Hampton et al. 2014), it was found that social media users are even less likely to express their opinions offline if they believe they differ from the majority opinion, speaking to the influence of socially-responding personas. It was also found that users want to make themselves seem more interesting and therefore shorten self-descriptions (Utz, Tanis, and Vermeulen 2012). Self-representation is also bound to time and place. In real life one must immediately respond to an interlocutor or opponent. In social networks, one has the option not to act immediately. Local binding is eliminated with social networks (Goffman 1959; Hogan 2010).

Presentation of self in terms of online media was theoretically addressed by (Hogan 2010). He contends self-representation is an increasingly frequent strategy in online participation. Following noted sociologist Erving Goffman's work (Goffman 1959), Hogan addresses digital 'exhibitions' and 'curators' where exhibitions are defined as status updates, listicles, or photos and the virtual curator creates the digital content. In setting the terms of self-presentation in theatrical terms, this work makes distinct that self-representation is the display of the ideal self, rather than a pattern of deception. Research on internet dating finds that the potential for self-representation is an attractive attribute of online activities (Lawson and Leck 2006). (Mehra, Kilduff, and Brass 2001) describe self-representation as self-monitoring, defined as the construction of a publically presented self for social interactions. The value of self-representation is supported by their findings looking at high and low self-monitoring (self-representing) by employees in a high technology firm. They find that high self-monitors are more likely to occupy preferential positions and have higher social network density than low self-monitors, measures of both the relative success of a self-representation strategy and common indicators of well-being (Huppert and So 2013). A contradicting study by (Ellison, Heino, and Gibbs 2006) considered an online dating environment in order to determine the extent of self-representation by users. Results of their interviews (n=34) indicate that the users who are more 'honest' in self-presentation have more success in dating. Nevertheless, all interviewees noted that in their online dating profiles they attempt to reveal themselves particularly positively, and have the same impression of the profile construction of other users. Across these studies, honesty in online representation is valued but ability and application of self-representation online has attractive socially-reinforced benefits.

6.1.2 Emotional Disclosure and Well-being on Facebook

Facebook's study on self-censorship, the typing then editing, deleting, or posting of statuses and comments from 3.9 million Facebook users, looks at how users alter their statements in quasi-public fora (Das and Kramer 2013). They found 71% of users self-censor in some way. Male users censor more than female, and Facebook posts are more frequently regulated than comments. They find that those with higher boundaries (estimated by the amount of regulations in place on the audience of the posting person) self-censor more, and theorize that the lack of control over an audience drives self-censorship. Perceived lack of control is generally understood to be a characteristic of neurotic personalities (DeNeve and Cooper 1998). Active self-censoring and its associated perceived lack of control can be understood as complementary to the findings of (Kross et al. 2013), who found that more time spent on Facebook is predictive of lower SWB, given the known relationship between low well-being and neuroticism.

Disclosure of emotional well-being online is different in real life (Qiu et al. 2012). In real life a person's feelings can often be guessed through facial expressions and body posture. Studies show that self-disclosure is generally more emphasized in real life. In (Qiu et al. 2012), it was discovered that users communicate their positive emotions more frequently via social posturing, finding that negative emotions in Facebook are hardly communicated. The intensity of positive emotion disclosure is linked to one’s extraversion or neuroticism levels as measured on the Five Factor model of (John, Donahue, and Kentle 1991). Propensity to disclose one’s emotional well-being is closely related to one’s personality (see Section 4.3), which is reliably measurable with online social media data.

Considering disclosure of personality and well-being it has been shown in this thesis and in literature that extraverts are linked to higher well-being and more positive emotional disclosure (DeNeve and Cooper 1998; Hall et al. 2013; Haslam, Whelan, and Bastian 2009; Yarkoni 2010). Neurotics have opposite tendencies. These personality types and disclosure patterns have unknown interaction effects with self-representation in online social networks. For an overview of this research, refer to Section 4.3. In accordance with this thesis’ findings and extant literature, the following hypotheses are established:

H1 *Extroversion is positively related to well-being*

H2 *Neuroticism is negatively related to well-being*

The hypotheses are key, as they substantiate the veracity of the data. If it is observed that the hypotheses cannot be rejected, then further assumptions about the underlying relationships between personality, well-being, and self-representation on Facebook can be made. Rejected hypotheses are then indicative of poor reporting from the platform, or overt self-representation.

A recent controversial study from Kramer and colleagues also employed emotional disclosure aspects, which can be understood as closely related to self-representation when considering the findings of (Hampton et al. 2014; Hampton et al. 2011; Utz, Tanis, and Vermeulen 2012). By altering the emotional content of friends’ statuses visible on the timeline, they found that the display of emotion is contagious (Kramer, Guillory, and Hancock 2014). Emotions in that study as well as this work are displayed via writing traits (as defined in Section 3.2.3). This study leads to the assumption that positive writing traits are linked to higher well-being and negative writing traits should indicate lower well-being, though this has not been definitively proven in literature. These findings lead to positing the following non-directional hypotheses in order to more fully investigate **RQ1**:

H3 *When well-being scores are high, more positive writing is used*

H4 *When well-being scores are low, more negative writing is used*

Establishing a relationship between well-being and writing traits allows us to extend the understanding of the relationship between personality, well-being, and online emotional disclosure. Rejecting H3 and H4 would indicate that no assessment between how well a person feels and their expression thereof on Facebook can be made, which is contrary to extant literature.

6.1.3 Detecting Personality and Well-being with Text Analytics

As reviewed in Section 2.2.2, LIWC is the tool in use in that it shows robustness to being used with short, informal text; it is available in multiple languages; and has the most extensive psychometric dictionary available to date. It has also been applied to similar social media studies looking at personality (i.e., (H. A. Schwartz et al. 2013)). These facts make it the most appropriate tool for the task of isolating personality from Facebook posts.

LIWC's premise is that it is structure and not context that matters. It argues that word function is more revealing than the words actually in use. Function words comprise approximately 55% of a given language and are difficult to manipulate (Tausczik and Pennebaker 2010). Function words can detect emotional states (Kramer et al. 2004; Kramer 2012), predict where they rank in social hierarchies and the quality of their relationships (Niederhoffer and Pennebaker 2002), along with their Five Factor Personality Model scores and happiness levels (C. Chung and Pennebaker 2014; Pennebaker, Mehl, and Niederhoffer 2003; Yarkoni 2010), as well as gender and age (H. A. Schwartz et al. 2013). LIWC has been applied to predict lying (Newman et al. 2003), and its output has proven to outperform humans and predict above random when detecting dishonest writing samples (Newman et al. 2003; Ott et al. 2011). Based on the findings of (Kramer 2010; Yarkoni 2010; H. A. Schwartz et al. 2013) two hypotheses on personality detection and writing style are grounded. Similarly to above, these hypotheses are not directional as the conversation has not been definitively settled in literature.

H5 *When extraversion scores are high, more positive writing is used*

H6 *When neuroticism scores are high, more negative writing is used*

The assumption is that personality is likewise identifiable in writing traits, concentrating on two traits well known to be associated with both positive and negative writing (see discussion in Section 4.2-4.3 on this relationship), and high and low well-being. Failing to reject these hypotheses indicates that it is possible to isolate personality traits in a one to one manner, as established in literature.

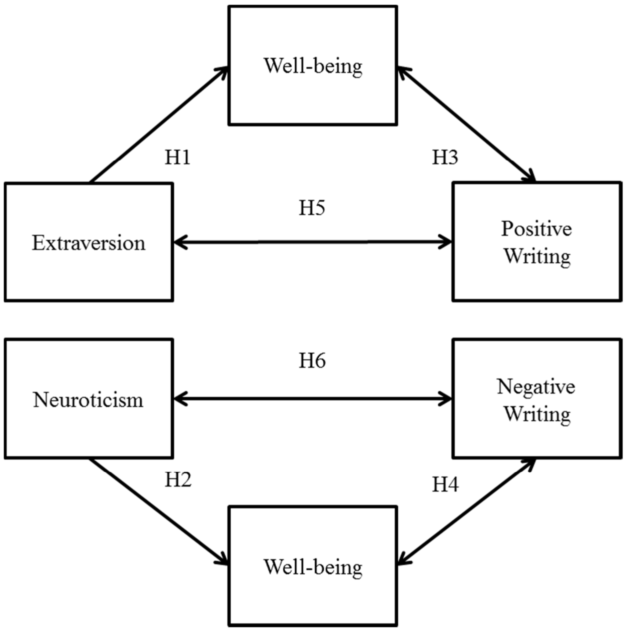


Figure 6.1: Relationship model considering directionality of personality, well-being, and profile text

These simultaneously considered hypotheses allow us to take a comprehensive view at the interactions between personality, well-being, and Facebook profiles (Figure 6.1) in accordance with **RQ2.4**. Whereas confirming H1 and H2 is necessary to validate the data, H3-H6 are useful in identifying if latent relationships exist as indicated in the data, or if there could be issues of self-representation present in data gathering from online social media profiles.

6.5 Methodology and Research Design

To facilitate the study, 509 AMT workers completed psychometric surveys via a Facebook application, from which 469 wholly-recorded questionnaires were returned. Whilst several approaches are available for discussion, including ex-post interviews with workers, this chapter concentrates on unobtrusive methods for the alignment of psychometrics and online social media persona. Psychometric surveys are a reliable and robust mechanism to establish personality, and can provide a necessary baseline of the person from which to diagnose self-representation. A selection of sentiment categories found to correlate with deception, personality, and confidence are then assessed to estimate individuals' propensities for self-representation in their social media persona (Buckels, Trapnell, and Paulhus 2014; Tausczik and Pennebaker 2010; Newman et al. 2003; Yarkoni 2010). While these indicators are unlikely to be the only psychometrics indicative of self-representation, but they are the most thoroughly researched and thus the most robust for this analysis.

In use for the establishment of personality is the instrument proposed by (John, Donahue, and Kentle 1991), the 44-item Big Five Inventory.[33] Human well-being and its expression are also of interest. To this extent, the Human Flourishing scale of (Huppert and So 2013) is employed in accordance with the discussion in Chapter 2.1.1. This 10-item scale established both SWB (Diener 1984b) and PWB (Waterman 1993), making it a valuable measure in the assessment of personal and emotional well-being.[34] In addition to the psychometric survey items is the 14-item online social media usage survey mechanism established in (Ewig 2011).[35] The question list and designation scheme is available in Appendix I. From this point on, all survey items will be referred to with their designated notation.

AMT has proven a reliable platform for conducting online experiments with a representative population (Berinsky, Huber, and Lenz 2012; Paolacci, Chandler, and Ipeirotis 2010; Ross et al. 2010). An initial screening question based on reading attentiveness was employed in order to minimize 'click-through' behavior (Berinsky, Huber, and Lenz 2012). Due to the question structure and number of questions, nine minutes was established as the minimum amount of time needed for completion. Workers who completed in less than nine minutes were excluded from the analysis, as well as those with unit or item non-responses, or otherwise incomplete items (Galesic and Bosnjak 2009; Bosnjak and Tuten 2001). The study was launched over a 24-hour period to accommodate differences in time zones.

A summarized privacy statement and informed consent document was presented on the entry page of the HIT (Human Intelligence Task), with a full privacy statement was available on request, detailing the uses of data and steps undertaken to guarantee privacy. Informed consent and privacy detailing are structured in accordance with the guidelines of the Association of Internet Researchers (Markham & Buchanan, 2012). As participants completed the survey, a PHP-based Facebook application simultaneously accessed their unique Facebook ID, and via Facebook's Open Graph API (application programming interface) accessed participants' Facebook timelines (Figure 2) for offline analysis. Payments of US$ 0.74 were issued at the end of the survey, equating to 1 cent per question. Participants' IDs were one-way hashed, with profile, survey, and worker payment being tied to the hashed ID. As the data is stored to disk, the hashing of IDs is necessary to maintain user anonymity.

[33] Big Five Inventory items are referred to as BF# in this chapter.
[34] Human Flourishing items are referred to as HF# in this chapter.
[35] Social media usage items are referred to as SM# in this chapter.

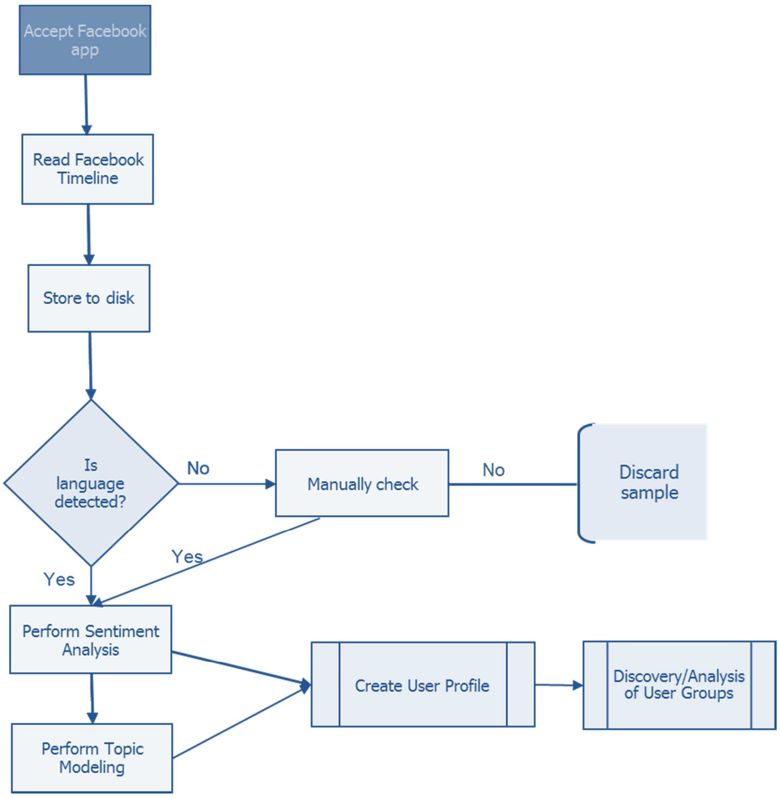


Figure 6.2: Workflow illustrating the steps to acquire, analyze, and interpret text data

Workers were given an option to opt out of the HIT at the stage where it linked to their Facebook profile or abandon the HIT at any other point. Privacy-aware users were able to hide their activities from the app. Regardless of users' privacy settings allowing timeline extraction or not, workers were paid with survey completion. The app extracted only posts, i.e., status updates, participants made to their timelines. Other post types such as shares, profile updates, etc. are excluded as they are not fully self-produced texts. This type of constraint can create first-order bias by potentially culling messages from the list of retrieved posts (González-Bailón et al. 2014). However presentation of the self, and mitigation of possible bias in self-presentation is under consideration; comments from other users are not immediately helpful. It is also an ethical grey zone to harvest the comments of participants' friends. As this study is not a network study, second order bias is not considered here (González-Bailón et al. 2014).

The JSON objects were retrieved from Facebook, parsed, and stored in flat files so that they could be imported into LIWC for sentiment retrieval. Procured data is stored initially in JSON objects (one per participant) and represents the entire timeline and basic information – this format mimics the Facebook representation of data, only without pagination. To analyze Facebook data, the data is partitioned with various granularities, i.e., per hashed ID or ID groups, and then temporally i.e., weekly, monthly, or the complete collection of posts for the entirety of the timeline. A complete description of the Social Observatory process is described in Chapter 5.2.1. Compiling the data in this manner allows execution of studies with LIWC at multiple granularities and time samples. The LIWC analysis is performed manually as LIWC does not facilitate automated invocation.

6.2.1 Statistical Modeling

Three statistical procedures are heavily utilized in this work, namely Spearman's ρ, logistic regression, and automatic linear modelling (SPSS version 22). Additionally, one secondary analysis required the application of an ANOVA (discussed in section 6.3). While linear relationships exist in the data, some cases are non-normally distributed. (R. L. Fowler 1987) notes that Spearman's ρ outperforms other correlation methods in cases of contaminated normal distributions, and is robust to Type III errors (correctly rejecting the null hypothesis for the wrong reason(s)). This justifies the use of ρ rather than Pearson's *r*, in spite of the fact *r* tests on true values rather than ranks (thus monotonic relationships). Spearman's ρ is calculated as:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} \tag{6.1}$$

For a sample of size *n*, with the *n* raw scores X_i, Y_i, raw scores are converted to ranks x_i, y_i, where $d_i = x_i - y_i$, is the difference between ranks.

Binomial logistic regression is appropriate for dichotomous dependent variables, such as those found in items [SM 4-7; 9, 11-14] and categorical or continuous independent variables (Rodrıguez, 2007). A binominal regression is formally described as:

$$\log \frac{p(x)}{a-p(x)} = \beta_0 + x * \beta \tag{6.2}$$

Where solving for *p* requires:

$$p(x; b, w) = \frac{e^{\beta_0 + x*\beta}}{1+e^{\beta_0+x*\beta}} = \frac{1}{1+e^{-(\beta_0+x*\beta)}} \tag{6.3}$$

Automatic linear modelling is employed for its facilities in automatic data preparation and handling. Regression in SPSS version 22 is ruled out as it is limited to step-wise methods only, cannot conduct an all-possible subset analysis (which is necessary here for exploratory reasons), and does not automatically identify and handle outliers. Automatic linear modelling is more robust against Type I and II errors in comparison, and can improve predictions by conducting a model ensemble (Yang 2013). The analysis utilizes the boosted, best-subset model consistent with data mining approaches, describes in Equations 6.4-6.9. SPSS 22 defines multiple imputation general linear regression as (IBM 2011a; IBM 2011b).

$$y_i = x'_i\beta + e_i \text{ with } e_i \tilde{} N\left(0, \frac{\sigma^2}{w_t}\right)$$

$$\text{Prior: } Pr(\beta, \log\sigma^2) \propto 1, \text{ or equivalently } Pr(\beta, \sigma^2) \propto 1/\sigma^2 \quad (6.4)$$

Using the complete cases (here, the survey data and results of the LIWC sentiment analysis), to fit the regression model. The assumption is that all redundant parameters (e.g., survey or LIWC categories) are removed. Denoting fitted parameters as $(\hat{\beta}, \hat{\sigma}^2)$ such that

$$\hat{\beta} = (X_c' F_c W_c X_c)^{-1} X_c' F_c W_c X_c$$

$$\hat{\sigma}^2 = \left(Y_c - X_c\hat{\beta}\right)' F_c W_c \left(Y_c - X_c\hat{\beta}\right)/(N_{obs} - p) \quad (6.5)$$

where $N_{obs} = \sum_{i \in obs(Y)} f_i$ is the number of complete cases, *p* is the number of parameters, and Y_c, X_c, F_c, W_c are the dependent vector, design matrix and frequency weight, regression weight matrix for complete cases.

The posterior distributions are:

$$\beta | \sigma^2, Y_c, X_c \,\tilde{}\, N\left(\hat{\beta}, (X_c' F_c W_c X_c)^{-1} \sigma^2\right)$$

$$\sigma^2 | Y_c, X_c \,\tilde{}\, (N_{obs} - p)^{\hat{\sigma}^2 / X^2_{N_{obs}-p}} \quad (6.6)$$

A is the upper triangular matrix of Cholesky decomposition $(X_c' F_c W_c X_c)^{-1} = A'A$ (6.7)

Drawing parameters from the posterior distributions, draw $(\sigma^*)^2$: defined as a random value *u* from $X^2_{N_{obs}-p}$, then $(\sigma^*)^2 = (N_{obs} - p)^{\hat{\sigma}^2/u}$. (6.8)

Draw β^*: draw *p* independent *N*(0,1) values to create a random vector *v*, then $\beta^* = \hat{\beta} + \sigma^* A' v$.

then imputing missing values. For *i* in *mis(Y)*, draw z_i from *N*(0,1); imputation is

$$y_i^* = x_i' \beta^* + \frac{\sigma^*}{\sqrt{w_i}} z_i \quad (6.9)$$

6.2.2 On Reliability and Method Biases

Surveys are prone to rater and item effects (Podsakoff, MacKenzie, and Podsakoff 2012) and online data is susceptible to context effects and sampling error (Sills & Song, 2002). The surveys in use are previously empirically validated and the data collection and processing found that 82% of the sample did not violate constraints suggested in (Podsakoff et al. 2003; Podsakoff, MacKenzie, and Podsakoff 2012). Chapter 4 shows that the Big Five Inventory and Human Flourishing (well-being) are reliably recorded in an online environment, mitigating context effects. The scales utilized had minimal social desirability and are balanced in positive and negative words (see Appendix I) in line with (Podsakoff et al. 2003; Podsakoff, MacKenzie, and Podsakoff 2012). The crowdworkers' results from these surveys indicate replication of (Huppert and So 2013; John, Donahue, and Kentle 1991; Ewig 2011), indicating reliable data.

The analyses suggest construct reliability and convergence, with the KMO measures for all constructs (personality, personal well-being, Facebook usage) ranging from 0.788 to 0.9 (Table 1). In the construct Facebook usage, a Principle Component Analysis (PCA) indicated that two traits, "Do other people present themselves differently in online and offline settings?" [SM10] (0.391) and "I can be more open online than in real life" [SM14E] (0.487) did not fulfil the KMO criterion of a 0.5 minimum value, and are therefore trimmed from the scale in accordance with (Podsakoff & Organ 1986). In each PCA analysis, Bartlett's test of sphericity was statistically significant ($p < .0005$), allowing rejection of the null hypotheses. This indicates that there are correlations between the variables, which are essential because if there are no correlations between variables, they cannot be factorized. Cronbach's α tests of internal consistency, a standard measure for this type of analysis, showed values ranging from 0.668 - 0.841 (Table 1). Generally speaking, an α above 0.6 is considerable acceptably consistent to be further researched (Lance, Butts, and Michels 2006).

Table 6.1: Measures of sampling adequacy and internal consistency

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		
Personality	Well-being	Facebook usage
0.648	0.900	0.788
Cronbach's α		
Personality	Well-being	Facebook usage
0.603	0.841	0.668

6.6 Results

Workers self-reported current locations in six geographic regions, with the bulk majority of workers reporting locations in the United States and India. Accordingly the largest language group was English with 285 timelines using predominately English. 73% of workers self-reported to be aged 35 or younger. Gender of the workers is evenly split between women and men, with one non-disclosure and one choice of 'Other.' 37% reported being unemployed and 57% completed at least a bachelor's degree. The boxplots of these results considering HFS can be found in Appendix IV.

Of the 285 English profiles, 282 have profiles with 50 or more words over the lifetime of the profiles (ranging from 2006-2014, with the average account opening in 2010). When considering the 285, the average word count per worker is 9,379; deleting these three profiles gives an average word count of 11,087. This signifies the magnitude of variance in the profiles. Table 6.2 illustrates some descriptive categories considering the average and the SD of the profiles, as well as the frequency of words with more than six letters, a measure of linguistic maturity (Tausczik and Pennebaker 2010). Again, emoticons and words per profile indicate a huge variance. Therefore, the following analyses are normalized for length unless otherwise stated. Only the 282 English profiles with more than 50 words are used unless otherwise noted as the profiles with 50 or fewer words do not have enough text for a proper analysis, and the other linguistic subgroups are likewise too small for meaningful statistics. The 50 word sensitivity threshold was determined via a repetitious data entry into LIWC; at the 50 word threshold there ceased to be significant differences in the percentages reported back from LIWC.

Table 6.2: Mean and SD per profile

Per Profile	**μ**	***σ***
Words Used	9379	24367
Emoticons	.05	.07
Unique Words	38	22
+6 Letter Words	16	6

There are some generally interesting results based on the calculation of Spearman's ρ, dealing with contact patterns and motivation of use outside of self-representation issues. Workers who use Facebook frequently also update their profiles frequently (r_s (337 = .292, p < .005) [SM 1/2], though those with a higher number of friends have a negative relationship with the frequency of logins (r_s (337 = -.314, p < .005) [SM 1/3]. A negative relationship also exists between number of the friends and the number of updates (r_s (337 = -.252, p < .005) [SM 2/3]. A worker with high well-being score has a positive significant relationship with a higher number of Facebook friends (r_s (337 = .112, p < .041) [HF/SM3], but a negative relationship with frequency of updates (r_s (337) = -.109, p < .047) [HF/SM2]. These results support, yet give a more nuanced understanding to the findings in (Kross et al., 2013) that Facebook usage predicts lowered SWB in young adults.

Family, and on and offline friends are a major interest areas for workers.[36] Workers who use Facebook to show what they know and can are less interested in contacting family than all other groups (on and offline friends, unknown people) (Exp(B) = 0.5, p = 0.071) [SM 9H/SM4]. Those who mainly like status updates are most likely to contact family members

[36] Results in this paragraph are the results of binomial regression.

(Exp(B) = 2.320, p =0.006) [SM 1D/SM4]. Workers who use Facebook in order to be recognized by others and are half as likely to have offline friends on Facebook as the rest of the population (Exp(B) = 0.550, p = 0.085), and are twice as likely to be interested in contacting family members on Facebook (Exp(B) = 1,989, p = 0,067) [SM 9C/4]. An exception here is those who want recognition and support from other users: they are half as likely to contact family members (Exp(B) = 0.406, p = 0.011) [SM 9E/4]. Men are less interested in maintaining contact with family on Facebook as women (Exp(B) = 0.393, p = 0.001) [SM4], and those who frequently like videos are twice as likely to use Facebook for contacting their family (Exp (B) = 2.502, p = 0.004) [SM5/4]. Workers whose profile picture does not show their face are half as likely to want to contact offline friends and are more interested in finding unknown online friends (Exp(B) = 0.413, p = 0.007) [SM 11F/4], as well as workers who have a stronger feeling of self-determination over what they show others (Exp(B) = 1.344, p = 0.033) [SM14B/4].

6.3.1 Identifying Self-Representation

Deceptive profiles as identified in (Newman et al. 2003) were assessed by first establishing the mean of the LIWC categories first person singular, motion, exclusion, and negative emotion. Two cut-offs were employed, by adding the first and second SD to the average. Those who employ above average negative emotion and motion words, and fewer exclusion words and less first-person singular are considered to display potential signs of lying. Fitting this description are 96 worker profiles, or 34 per cent of this sample. This is line with the findings of (Caspi and Gorsky 2006), who found about a third of Facebook users regularly lie in their Facebook interactions. These profiles are demarcated in order to use them as a control element.

If H1 and H2 are confirmed, the assumptions are that H3 - H6 should also be confirmed; otherwise, issues of self-representation in the data are likely evident in the data. For the one-tailed hypotheses of a positive relationship existing between well-being and extraversion and a negative relationship existing between neuroticism and well-being [H1/2], both hypotheses are strongly confirmed ([$r_s(282) = .357$ $p < .0005$] / [$r_s(282) = -.263$ $p < .0005$]).

Table 6.3 shows the further breakdown of H3 and H4 from “writing traits” into their respective LIWC categories and well-being. H5 and H6 are likewise expanded to assess personality and the related LIWC categories (Table 6.4). Considering well-being and writing traits, only H3c is confirmed, namely there is a relationship between that of well-being and optimism. Those who are flourishing will accurately portray their propensity to feel optimistic in their writing, though nothing else, where those who have lower emotional well-being seem to self-represent their traditionally negative views outside of their Facebook information.

Table 6.3: Summary: Hypotheses on the relationships between happiness and LIWC categories

		ρ	P	✓/≈/¬
H3	When well-being scores are high, more positive writing is used	-	-	≈
H3a	When well-being scores are high, more (written) positive emotion is used	.102	.088	¬
H3b	When well-being scores are high, more positive feelings are used	.030	.612	¬
H3c	When well-being scores are high, more optimism is used	.144*	.015	✓
H4	When well-being scores are low, and negative writing traits	-	-	¬
H4a	When well-being scores are low, more (written) negative emotion is used	.016	.785	¬
H4b	When well-being scores are low, more anxiety is used	-.035	.557	¬
H4c	When well-being scores are low, more anger is used	.029	.625	¬
H4d	When well-being scores are low, more sadness is used	-.025	.682	¬

Personality and writing traits have likewise one significant relationship, neurotic personality types and expressed anxiety on Facebook (Table 6.4). This indicates that self-representation is likely to be higher with those who self-identify as extraverts, whereas neurotic personality types do leave some digital indicators of their personality.

Table 6.4: Hypotheses on the relationships between personality and LIWC categories

		ρ	p	✓/≈/¬
H5	When extraversion scores are high, more positive writing is used	-	-	¬
H5a	When extraversion scores are high, more (written) positive emotion is used	-.019	.751	¬
H5b	When extraversion scores are high, more positive feelings are used	-.031	.598	¬
H5c	When extraversion scores are high, more optimism is used	-.016	.795	¬
H6	When neuroticism is high, more negative writing is used	-	-	≈
H6a	When neuroticism is high, more (written) negative emotion is used	.069	.402	¬
H6b	When neuroticism is high, more anxiety is used	.120*	.043	✓
H6c	When neuroticism is high, more anger is used	.061	.307	¬
H6d	When neuroticism is high, more sadness is used	.050	.398	¬

As H1 and H2 are confirmed, whereas only H3c and H6b are confirmed of the remaining 18, it indicates that workers have (either on purpose or inadvertently) systematically self-represented themselves on Facebook. When statistically controlling for deceptive profiles, the weak significances of H4b and H5c disappear. This could be a confirmation that deception and self-representation are conceptually different, supporting the framework of (Hogan, 2010). Having identified that the data is reliable, it is clear that relationships between personality, well-being and text are undermined by the online medium. This necessitates controlling for participant-induced bias in research designs where the veracity of self-produced texts is necessary for interpretation.

Workers generally communicate their positive emotions more frequently (an average of 4.25% of all text), where negative emotions in Facebook are hardly communicated (1.2% of all data), regardless of Five Factor personality type and in line with the results of (Qiu et al. 2012). As 60% more words of the LIWC dictionary are associated with negative sentiment, the social posturing aspects are clear. This chapter identifies “displays of positive emotion” and “hiding negative emotion” as forms of a self-representation bias. This could also be a contributing factor to the findings of (Kramer, Guillory, and Hancock 2014).

The analysis also considered expressed confidence as a measure of self-representation. This is measured by the mean frequency in usage of first person singular and third person plural;

where people that are more confident use "I" words less than "We" words (Pennebaker, Mehl, and Niederhoffer 2003). Here the demographic groups established in the survey are tested with an ANOVA (Figure 6.3) and found a significant difference in gender (Gender $F(2{,}279) = 11.893$, $p < .0005$; Wilks' $\Lambda = .921$; partial $\eta^2 = .079$). The findings cannot reject a difference between third person plural between men and women (First Person Plural (We) $F(1{,}280) = .643$, $p = .423$; partial $\eta^2 = .002$), whereas first person singular has a significant difference in gendered usage (First Person Singular (I) $F(1{,}280) = 23.405$, $p < .0005$; partial $\eta^2 = .077$). There was homogeneity of variance-covariance matrices, as assessed by Box's test of equality of covariance matrices ($p = .002$). This supports emerging findings[37] that women express less confidence than men do, and thereby does not support overt self-representation specific to online social networks. This is an interesting finding because whereas there are no gender differences found in the rejected hypotheses indicating self-representation, males are significantly more likely to truthfully present their confidence in their online personas. Based on the findings of (Das and Kramer 2013), that men self-censor more, this is an unexpected finding. There is no relationship between deceptive profiles and confidence.

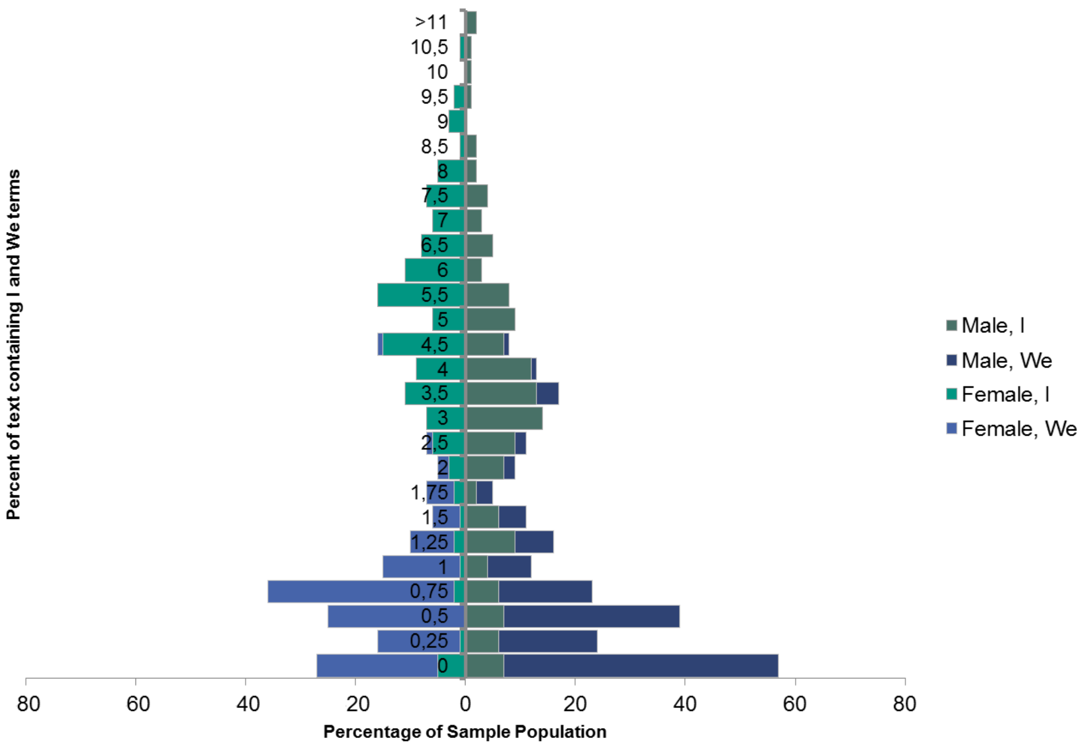


Figure 6.3: Gendered usage of confident statements on Facebook profiles

In a response to **RQ 2.4**, self-representation is present and identifiable. Its contours are evident in self-produced text. Specifically the masking of personality and well-being, as well as the masking of negative emotion are indicative of self-representation (**RQ 2.3**). Deceptive

[37] http://www.theatlantic.com/features/archive/2014/04/the-confidence-gap/359815/

tendencies in self-produced text are also identifiable, and deception is conceptually different from self-representation in online social networks.

6.3.2 Personality as a Tool for Mitigating Self-representation

Workers responses to the Five Factor model and Human Flourishing items proved to be indicative of self-representation when compared to their self-produced text. Applying the data mining technique referred to in Section 6.2.1 (Equations 6.4-6.9), 136 variables[38] of survey responses and sentiment categories on each of the five personality traits of the Five Factor model (John, Donahue, and Kentle 1991) are regressed, using the created ‘deception’ variable as a control element. The approach creates meritorious model fits averaging 74.6% accuracy as presented in Table 6.5, without overt signs of overfitting. The multivariate models are statistically significant for each personality trait, with some overlap of the variables predicting the traits. Considering sizeable correlations between predictor groups, the unique variance explained by each of the variables indexed by the squared semipartial correlations is low. In no case is Cook’s Distance larger than one; outliers were accordingly handled within the data rather than trimmed. The coming section is a short discussion of the predictors of each trait, with predictors grouped by measurement instrument then listed by weight. In order to constrain the number of variables, the ten items’ strongest relationships’ significant at the (p < .001) level per trait are reported.

Table 6.5: Prediction accuracy per model on Five Factor Personality traits, boosted (10 component models) using best-subsets

Trait Name	Reference Model	Ensemble	s^2
Openness	78.5	77.3	1.2
Conscientiousness	69.4	64.3	5.1
Extraversion	77.8	69.5	8.3
Agreeableness	71.4	71.0	0.4
Neuroticism	75.9	68.9	7.0
Average	*74.6*	*70.2*	*4.4*

Openness has the highest prediction accuracy of 78.5%, and is a very stable prediction given the low difference indicates that the prediction is relatively stable. Highly significant are the survey categories meaning [HF 4], self-esteem [HF 9], engagement [HF3], competence [HF 1],

[38] Punctuation and the corresponding Big Five traits are excluded from this regression. A component table is available in Appendix 2.

optimism [HF 5], positive emotion [HF 6], and resilience [HF 9]; the country of origin of the worker; and the LIWC sentiment category Feelings.

With the lowest prediction accuracy (69.4%) and a medium model difference (5.1%), ***Conscientiousness*** must be considered less reliable. The LIWC sentiment categories, Friends, Down, and Fillers; survey responses ‘a profile picture that is not obviously me’ [SM11F], number of friends [SM3], ‘I understand quickly how others perceive me’ [SM 14A], assent to ‘People should present themselves on online social networks as the same person as they are offline’ [SM 8], and using Facebook to give and get information [SM 9K], and the survey measurement resilience [HF 9] and positive relationships [HF 7] are the most relevant predictors.

Extraversion with 77.8% accuracy and the largest difference of 8.3% is related to the survey items competence [HF 1], self-esteem [HF9], meaning [HF 4], optimism [HF 5], positive emotion [HF 6], vitality [HF 10], and resilience [HF 9]; country of origin; and the survey responses ‘I understand quickly how I am perceived by others’ [SM 14A] and managing Facebook profiles with displays of albums [SM 11G].

Agreeableness has the lowest deviation (0.4%) and an accuracy of 71.4%, indicating high reliability. Highly significant are the survey items resilience [HF 8], meaning [HF 4], self-esteem [9], and competence [HF 1]; country of origin; the sentiment categories Friends, Inhibition, Feelings, and Assent; and declination of ‘I can be who or what I want on my Profile page’ [SM 14D].

Neuroticism has a high deviation between models (7%), but a good performance (75.9 % accuracy). As established in Section 4.2.1 it is imperative that neuroticism have high prediction accuracy, as it is the trait with the highest predictor weight in well-being assessment. The most significant survey items are resilience [HF 8], self-esteem [HF 9], emotional stability [HF 2], vitality [HF 10], and optimism [HF 5]; using Facebook to spy on others [SM 9D], managing presentation of self with pictures not of them [SM 11F], using Facebook to observe other people [SM 9F], and liking videos on Facebook [SM 5]. Additionally, the LIWC sentiment category Feelings is highly significant.

As the use of text, and not survey items, would be the only available data ‘in the wild,’ only on data that would be available from Facebook profiles to define the relationships between LIWC and the personality is assessed. The sub analysis shows that topical discussions have high prediction value for the Five Factor model (Table 6.6). Highly significant for openness are the sentiment categories Sports, Religion, Feelings, Music, Fillers, and TV, where Sports, Music, Fillers and TV have a positive association with openness; Feelings has a negative association; and Religion has an inverted U-shaped relationship with very low and high openness scores have a positive association, but mid-range having a negative association. Conscientiousness displays that Religion, Friends, TV, Inhibition, and Music are positively related, and Fillers is negatively related. Extraversion is positively related to Inhibition and TV, and negatively related to Friends, Sports, and Down. Agreeableness' highly significant sentiment categories are negative relationships with Inhibition and Death, and a positive relationship with Friends. The final trait, neuroticism finds Religion, Friends, TV, Inhibition, and Music being positively related and Fillers being negatively related.

Table 6.6: Five Factor Model mapped to positive and negative relationships of LIWC sentiment categories with high predictor strength ($p < .001$)

Openness		**Conscientiousness**		**Extraversion**		**Agreeableness**		**Neuroticism**	
+	-	+	-	+	-	+	-	+	-
Sports	Feelings	Religion	Fillers	Inhibition	Friends	Friends	Inhibition	Religion	Fillers
Religion	Religion	Friends		TV	Sports		Death	Friends	
Music		TV			Down			TV	
Fillers		Inhibition						Inhibition	
TV		Music						Music	

While surprising at first glance, when the medium of data is considered the findings are less surprising. Facebook is a medium to exchange news and ideas, and while more reflective in nature and practice than Twitter (Dodds et al. 2011), is still essentially used as a short information service to connect people (Hampton et al. 2011; R. E. Wilson, Gosling, and Graham 2012). Several sentiment categories dominate the results; specifically inhibition is very common, suggesting that workers (consciously or not) are in fact utilizing vocabulary of inhibition on their Facebook profiles. This could be further indicative of self-representation.

Thus established, researchers may now use these patterns to identify personality without the need for costly, traditional survey methods. Utilizing a similar method as employed to define deception as in the Analysis section can reveal the tendency of the profile, thus allowing the researcher to build a single variable from which to create a dummy. Said dummy can be used as a control factor in the analysis of online social media data. In short, mitigation of self-representation allows for mitigated researcher bias in the translation from the way that people think and behave to their digital traces of thoughts and behaviors.

6.7 Discussion and Limitations

The key findings of this work are that self-representation in online social media is an identifiable phenomenon, that self-representation can be isolated, and a number of indicators can be used to do so (**RQ 2.4**). Personality in particular can be used a supporting factor in mitigating self-representation, further supporting its importance to TSR frameworks and applications. Identifying self-representation contributes a method for social researchers to verify psychometric baselines of subjects by mitigating the effects of socially responding personas in online social media data. Moreover, it opens an interesting discussion on the impact of self-representation on social media analyses, both from the perspective of the researcher validating social models, and the subject considering their intention of such behaviors. The text samples were generated in a way which did not induce measurement errors in accordance with (González-Bailón et al. 2014; Ruths and Pfeffer 2014). Whilst profiles indicative of deception are identified in the text-based sample, the control measures noted above mitigated this. Profiles indicative of deception are isolated, and used as a control item.

Self-representation was identified in a number of indicators (**RQ 2.4**). While the survey-only results show a replication of literature, the survey to text results cannot replicate the findings that extraversion is a predictor of well-being, and neuroticism has a negative relationship with well-being. Positive affectivity and withdrawn negative emotions are identifiable across all workers' profiles. One value contribution is the finding that withdrawn negative affect is a particularly indicative of self-representation. This further supports the use of a multi-dimensional sentiment analysis rather than a focus on positive and negative emotion for assessing communal well-being (**RQ 1.1**). Confidence can be identified and follows expected patterns across genders. Male participants appear more confident in their written profiles than females. As this is a finding in emergent literature, this cannot be understood as an overt measure of self-representation.

Given the highly clustered, trivial nature of the sentiment-based predictors, a tempting statement is that the data is not appropriate for the task. However, discernable patterns are present. Especially the strength of inhibition in four of five of the Five Factor model suggests that the participants display reticence about showing their actual personalities in their Facebook profiles. Moreover, given the platform, the topics discussed are a reasonable (albeit, surprising) output. The topical basis of the other predictors conceptual themes of workers' discussions, and neatly creates psychological profiles that links online and offline personality. In future TSR applications, stakeholders and researchers are able to control for these categories and their positive or negative relationships in data preparation or as a control factor in the calculation, e.g., as a dummy variable in regression models.

This study is not without fault. Firstly, the applied method is an estimation and not a revealed method, as is more common to the Social Observatory. This leaves room for errors. A limitation is the sample size, which disallows larger statements about subgroups as the non-English samples are too small for meaningful statistics. Another drawback is that the results are tailored to Facebook – the findings of this study are unlikely to generalize to professional networking, microblogs, or visual media sites. A known issue of Natural Language Processing is that the state of the art tools are unable to capably handle sarcasm and irony (Tsur and Rappoport 2010), which has unknown effects across the lifespan of a Facebook timeline. A concluding remark on limitations is related to privacy. While the study obtained informed consent of its workers, the open question remains if workers truly understood the amount of information that was being given in the HIT.

6.8 Summary and Implications

The stated aims of this chapter are twofold: establishing the relationship between offline and online personalities, and mitigating of biases in surveys and in publically sourced data. In accomplishing these goals, this thesis creates a generally applicable method in support of the Social Observatory and its stated aim to unobtrusively analyze social phenomena like well-being or other social indicators (**RQ 2.4**). Such a method is impactful in both research arenas and commercial domains, in that it allows the study designer to approximate participant baselines without highly intrusive mechanisms. In a systematic manner, this research detailed the experimental design, data collection, and analysis. Common method biases are addressed and appropriately eliminated when identified. The method allows for replication by careful detailing of the steps and processing of data.

A strength of this chapter is its consideration and application of the findings from recent cyber psychology literature to identify and isolate established elements of well-being and personality. A major contribution is addressing method biases in the harvesting and analysis of social media data. This research utilizes the entire data stream per profile, mitigating first order bias. With personality and well-being validated, and a sentiment analysis performed on the lifespan of a user’s Facebook timeline, the propensity of a user to portray themselves in opposition to their truthful, psychological baseline is revealed. It also names common markers of the phenomena of self-representation based on simple sentiment categories and psychometrics that allows researchers to mitigate its effects in future TSR applications.

Natural extensions of this research are closely linked to its limitations. Cross-platform analysis of the same user for their various public profiles would give future work a more nuanced view in the ways that social media users self-represent in difference audiences. Such a work would fill research gaps in ‘best’ platform usage for information disbursement, creation, and influence, as well as network impact. A network analysis with a textured understanding of how users cluster and complement within a network would be a good area of future research.

Researchers can apply this method to their analyses of publically sourced data in order to mitigate the effects of various phenomena, including trolling, social desirability, and acquiescent behaviors (e.g., the spiral of silence). Such an approach has diverse applications in that it allows for a new, accurate measurement system from which to deduce from publically accessible text onto the general population. With self-representation identified, a valid measurement of psychometrics without necessitating expensive survey methods is created.

Chapter VII Applied Institutional Well-being: A Case Study on KIT

"The care of human life and happiness, and not their destruction, is the first and only legitimate object of good government."

Thomas Jefferson (1809)

The dividing line between offline and online communities is increasingly intertwined. Cases where physical presence was assumed to be a foremost asset are becoming less common. The clearest example is the 'brick and mortar' of the world's top universities slowly transitioning to MOOCs. Such a transition impacts innumerable processes, giving unprecedented space to innovate and improve. One such area prime for improvement is institutional quality and satisfaction rankings at universities. Current metrics share the same characteristics, namely that they are externally audited, time-lagged macro-assessments, requiring little to no participation from stakeholders. These problems mean that current rankings leave a lot to be desired in terms of transparency, engagement, and time-sensitive integration. Current ranking efforts are deficient. As succinctly put by the European University Association's working group on university rankings in their report 'Rankings in Institutional Strategies and Processes':

> "*Ultimately, to overcome problems associated with inappropriate indicators used by rankings, should there be an international common dataset on higher education which would facilitate greater and more meaningful comparability? As challenging as it may be to find consensus on such a dataset, it might be worth exploring the possibility* (Hazelkorn, Loukkola, and Zhang 2014, 50)."

The urgency and merit of this assessment is due to the public nature of university rankings: students as well as public funding bodies take note of such information, and can take make decisions on enrollment, transferring, and grant allocation based on it (Hazelkorn, Loukkola, and Zhang 2014). Especially considering the perspective of university stakeholders, a novel approach to rank the performance of universities would be to assess the university community's subjective opinion(s) of its campus and its programs, aggregating based on quality and selected social indicators like communal well-being. In terms of TSR, such a platform would establish a more granular and sensitive feedback system for stakeholders (i.e., university administration, students, faculties) to assess and respond to university performance.

In response to this a Social Observatory is employed to find, analyze, and report socially-sourced indicators on university quality and satisfaction. This is well in-line with the proposed TSR framework of Chapter 3: needed is a system that is conscious of the person, and the environment that person exists in, to evaluate (and eventually raise) well-being overall. The Social Observatory procured data from popularly used public Facebook pages surrounding the Karlsruhe Institute of Technology (KIT), for a tool that is near to real time and sensitive to concerns of both privacy and the desire to participate in decision making. The Social Observatory focuses specifically on the extraction and analysis of well-being as an alt-metric, in line with efforts to consider stakeholder well-being in policy and service applications (see Chapter 2.2.1 for an overview of current well-being indices and Chapter 3 for TSR).

Considering fast-paced online communities there is an institutional interest in knowing if, and which, events have significant effects on the way the community interacts and expresses itself (online), and if there are sentiment changes over longer time periods. These are isolated and extracted as measures of communal happiness and satisfaction. This chapter is the extension of the work in progress paper (Lindner et al. 2015), which presented a subset of the data analysis as a proof of concept work at the ACM CHI conference. Section 7.1 justifies the design made in the implementation choices and gives the descriptive attributes of the KIT Facebook network. Section 7.2 reviews the macro, meso and micro attributes of communal discourse across the KIT Facebook network. Section 7.3 discusses and contextualizes the findings, and Section 7.4 addresses limitations and concludes the chapter.

7.1 Study Design and Approach

To address research questions several steps must first be taken. The data must be prepped, the sentiment scores established, and then the sentiment scores must be audited for self-presentation. Only then is the data sufficiently prepared for the assessment of communal well-being. The coming sections address and discuss the design aspects behind TSR requirements for a Social Observatory based on Facebook data.

7.2 Macro, Meso, and Micro Granularities of BeWell@KIT

The first assumption to be addressed it the use of Facebook as opposed to Twitter. The KIT study database features an average text length of 33.96, mainly German, words. If the average German word length is estimated as 5.7[39] this would exclude 33.57 characters of the average message or otherwise force unnatural brevity or improper spellings. The fraction of posts and comments in this procured dataset containing more than 160 letters (28 words on average) represents 80.1% of the corpus, reflecting 39.86% of all comments and posts being longer than

39 http://www.duden.de/sprachwissen/sprachratgeber/durchschnittliche-laenge-eines-deutschen-wortes. Last Accessed: 10 March 2015.

Twitter's restriction. Using Twitter would certainly result in drastically shorter text submissions and consequently in a loss of more complicated, reflective statements. There is an additional restriction of Twitter that lends an unknown bias, namely that Twitter grants between 1-10% of the data available from the first request date in a given query (González-Bailón et al. 2014; Ruths and Pfeffer 2014; Russell 2013), compared to the full Timeline of the Facebook extraction. Most importantly, the choice of platform should consider the prevalence of the specific use case on the various networks. For KIT, Facebook usage outranked all other Social Media in this area for both university-generated and student-generated content, which is in line with the fact that Facebook has an 82% market reach of Germany, whereas Twitter has approximately 20%.[40]

In order to gain a more granular understanding of how the KIT relates and interacts within its online community, the baseline of discourse and latent emotive value must be established. This created the design choice of focusing on the years 2011-2014; while some pages were open longer than this, all pages included in the study were open from 2011 onwards (though sometimes inactive). Four granularities are investigated: post-comments splits, page group splits, administration-faculty splits, and individual posts and comments. The details of how the page splits are made are addressed in the chapter before the corresponding analysis is introduced. From this baseline it is possible to see what, if any, spikes and dips appear. Estimation the reasons for these spikes and dips can either be either temporal (event-based), well-being related (psychometrics) or both. Accordingly this chapter describes the KIT Facebook community, establishing the attributes which make up the communal discourse. From this point, the data is inspected for sentiment-based irregularities that could signify major community events (emotional or otherwise).

7.2.1 Macro Attributes of the KIT Facebook Network

The raw data from the database is first filtered based on based on post type, then aggregated to represent groups of the university (discussed in more detail in Section 7.2.2), run through LIWC and finally mapped and assessed. All data is normalized per granularity assessment to assure common baselines. From a corpus of 2,032,323 words, 1,806,232 were from posts and 226,091 were from comments. The social graph was rebuilt by weighting resources on an interaction basis (Figure 7.1). This graph reflects direct interactions considering activity on a page such as posting, liking, tagging or sharing of and commenting on content. Per contra, indirect relationships are generated when common third parties execute actions on both Facebook pages' timelines or, vice versa, a third party has an activity appear on its timeline by both pages. The resulting graph depicts the relative contribution of each page to the total data magnitude by sizing the nodes accordingly. Similar to the graph discussed in Section 5.4, positioning near to the center indicates that the page is well integrated into the community as a

[40] http://www.statista.com/statistics/280176/penetration-rate-of-social-media-sites-in-germany/. Last Accessed: 10 March 2015.

whole, whereas pages far on the outside have low interactions with other pages and audience members (e.g. KIT Career Service). Furthermore, edge thickness indicates stronger network ties based on the observed interaction frequency. The main KIT page acts as the central node in this graph. Figure 7.1 shows the most highly weighted edges, meaning that a node in the figure has a high centrality, or relationship, with the main page of KIT. The most central faculties are Economics & Industrial Engineering, Computer Science and Mechanical & Chemical Engineering. Regarding social aspects, KIT's German and English pages of the Germany-wide 'Spotted' dating pages are also strongly linked and quite central to the KIT Facebook network.

Table 7.1 gives further descriptive attributes of the dataset. In line with Chapter 5, likes far outnumber posts and comments, and posts outnumber comments. That posts outnumber comments in this use case is a surprising characteristic as most official pages only permit administrators to post on the timeline; constituent participation is restricted to commenting on those posts.

Table 7.1. Sum of values of all pages in KIT Facebook network considering possible interactions of the pages and audiences

Page Likes	Status Updates	Wall Posts	Comments	Likes on Posts	Resources Posted	Resources Liked
101,772	26,259	4,284	16,079	179,721	8,817	45,241

Self-representation, as defined as the misrepresentation of self on online social media in Section 6.3.1, represents the last data preparation step of the KIT Facebook database. Section 6.3 suggests isolating the LIWC correlates of the posts and comment's Five Factory Personality tendencies to identify self-representation. In order to assess if pages can be identified as applying self-representation, posts and comments that are over two SDs outside of the respected LIWC category are identified (a similar process to identifying deception in online social media from Chapter 6.3.1). Considering the outer boundaries of two SDs outside of the mean, no pages' posts or comments were identified as displaying the profiles of Openness, Conscientiousness, or Agreeableness. The posts of the Library were identified as displaying possible Extraversion traits, and the posts of the KIT Music page was identified to display possible Neuroticism traits (Table 7.2). The posts are identified as tending towards showing self-representation but not fully indicative of self-representation for two reasons:

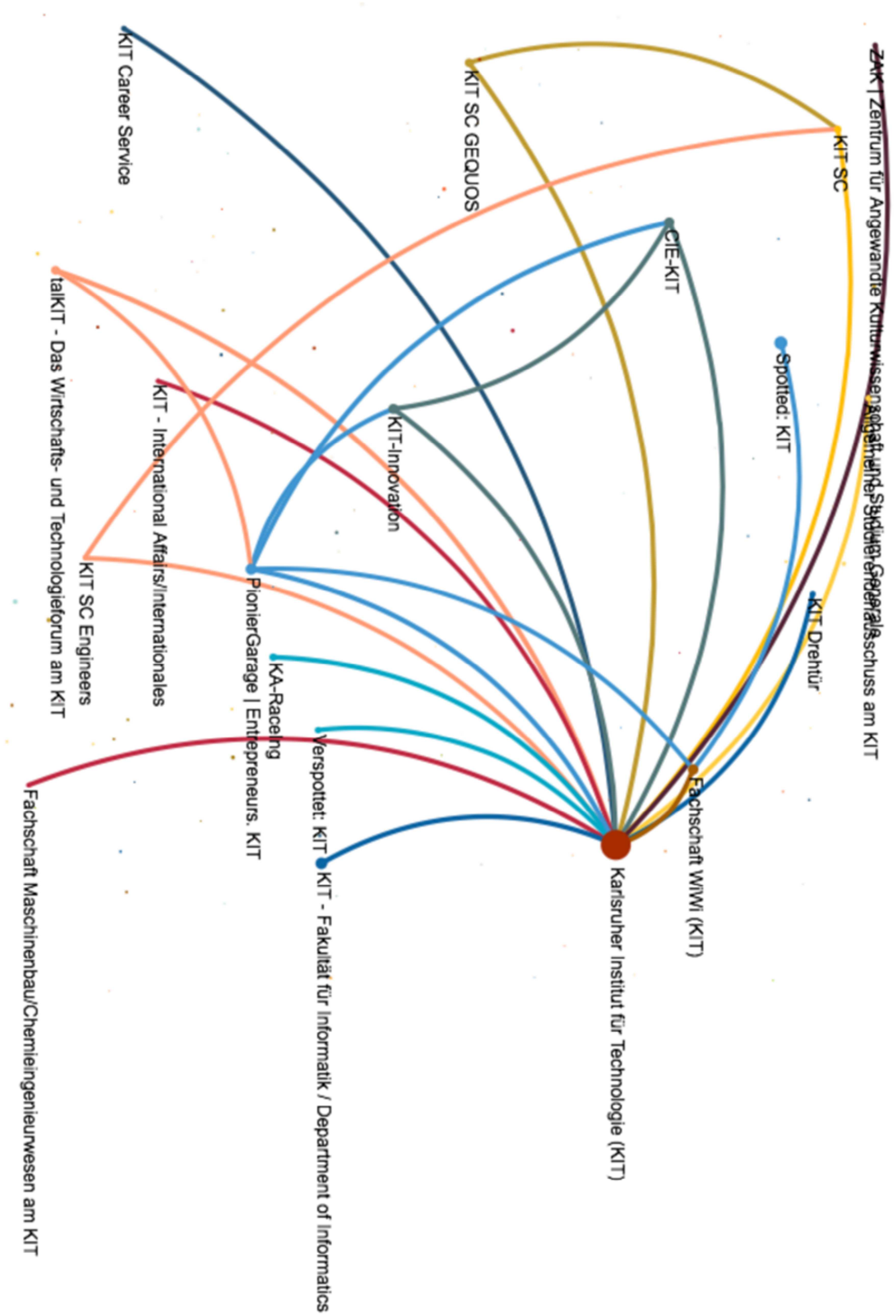


Figure 7.1 Network graph of the KIT pages considering all interactions, depicting most important nodes and edges

1) Each page is a majority but not 100% match to the trait characteristics defined in Chapter 6.3.1.
2) This method is an estimation method and not a revealed method.

The previous points require that the data of these pages be put to consideration but not that it be extracted from the dataset. These posting groups are therefore treated to control elements (verification via dummy testing) in order to verify that the analyses are valid and reliable, as well as similar to the actual posters in intent. They are included in all future analyses.

Table 7.2. Relationships of LIWC sentiment categories with high predictor strength ($p < .001$) of self-representation where green signifies above the second SD and red signifies below the second SD

Extraversion (Library)						Neuroticism (Music)					
Positive			Negative			Positive			Negative		
LIWC Name	2nd SD	Page Value	LIWC Name	2nd SD	Page Value	LIWC Name	2nd SD	Page Value	LIWC Name	2nd SD	Page Value
Inhib-ition	0.41	0.90	Friend	0.08	0.05	Rel.	0.21	0.34	Fillers	0.00	0.00
TV	0.43	0.04	Sports	0.00	0.04	Friend	0.23	0.25			
			Down	0.05	0.04	TV	0.43	0.94			
						Inhib-ition	0.41	0.10			
						Music	1.22	4.43			

KIT's communal discourse has a cyclic pattern that matches recurring semester cycles: The start of semester, mid-semester, exam weeks and semester holidays. The intensity of interactions also follows this pattern closely, as approximately 66% of interaction occurs inside of the semester (Table 7.3). It must be noted that as this study ranges from 2011-2014 the exact start and end dates of semesters are approximated by taking the mean of the official semester calendar.

Table 7.3. Semester cycles of the KIT Facebook network

Semester Intervals at KIT		
Start of Semester: %WC: 16%	Winter: 10/7-10/31	Summer: 04/07-04/31
Mid-Semester: %WC: 50%	Winter: 11/01-01/24	Summer: 05/01-07/09
Exam Weeks: %WC: 18%	Winter: 01/25-03/13	Summer: 07/10-08/14
Holidays: %WC: 16%	Winter: 03/14-04/06	Summer: 08/15-10/06

WC%: Percentage of Total Word Count

7.2.2 A Meso-assessment of KIT's Discourse Baseline

A group representation is the creation of supra-groups based on commonalities (e.g., administrators and students, faculties, student groups) used to assess the KIT community as a more realistic replication. Regarding group partitioning, two approaches are executed. First, all the 140 available pages are assigned to one of 12 page categories in order to facilitate analyses of the university's Facebook community. The naming of the groups is guided by the KIT website where possible to assure a realistic assessment in reconstructing discourse. In the case of KIT affiliated but not KIT sponsored groups, the most general common name is used. The names of the groups are KIT (official presence), Library, Schools, Departments and Institutes, Student Clubs, University Clubs, Sports Teams, Innovation and Development, Politics, Career, Music, and Social. An overview of this subdivision, along with the names of all available pages, is available in Appendix V. It must be noted that during the course of the study five pages closed and were duly excluded from the analysis; pages with less than 50 words over the four years of assessment are likewise excluded (as established in Chapter 6). These groups can be then further assessed considering if they are run by administrators or students. Splitting the data into these subgroups aims to reproduce an accurate picture of the community, by taking interactions and communal diversities within into account. At the same time it reflects an opportunity to extend the partitioning types discussed in the preceding paragraph.

A nearest neighbors calculation based on Euclidean distance over 64 LIWC categories is performed, similar to Chapter 5; Equation 5.1. This Chapter likewise measures k=5 neighbors for each of the 95,040 possible segment combinations (*x* and *y*), the squared difference scores of the identical LIWC category are added over a 64-dimensional plane. The measure of distance results when taking the square root of this sum. Higher distance scores reflect higher dissimilarity of two page categories. The results of the nearest neighbors analysis are available in Appendix VI. The absolute range of the 24 segments is highly clustered (10.39 – 11.22), indicating that some elements of hubness may be at work due to the high dimensionality of the data (Radovanovic, Nanopoulos, and Ivanovic 2010). However, some distinct patterns are still revealed. The most immediate revelation is that comments are quite diverse in comparison to posts. Posts tend to be most similar to other posts; in only three cases do posts have comments as one of their nearest neighbors. The most notable exception here is for the posts of Social pages, which tend to be more similar to comments. This could be a reflection of the fact that Social pages tend to be managed by students and not university administrators. The same is not true for comments, which average between 2-3 post-based neighbors. Music-related Facebook pages are the only case where the post-comment combination is placed at k=1.The next instance where a posts-comment combination overlap is within the Faculties, where k=4. Interestingly, this approach replicates the mapped interaction graph well (Figure 7.1); the most similar categories also make up the more interactive individual pages of the Facebook network.

The KIT network expresses itself as very inclusively. It is interesting to note that the use of exclusion, while minimal, spikes in comments and dips in posts (Figure 7.2). Music posts and comments have an observable dip in the use of both, indicating that these pages' discourse tends to be outside of including or excluding audiences. An observation of the data found that the Music pages tend to be more informative, declarative statements. This discrepancy in usage could be due to this aspect. It could also be an aspect of self-representation as discussed earlier.

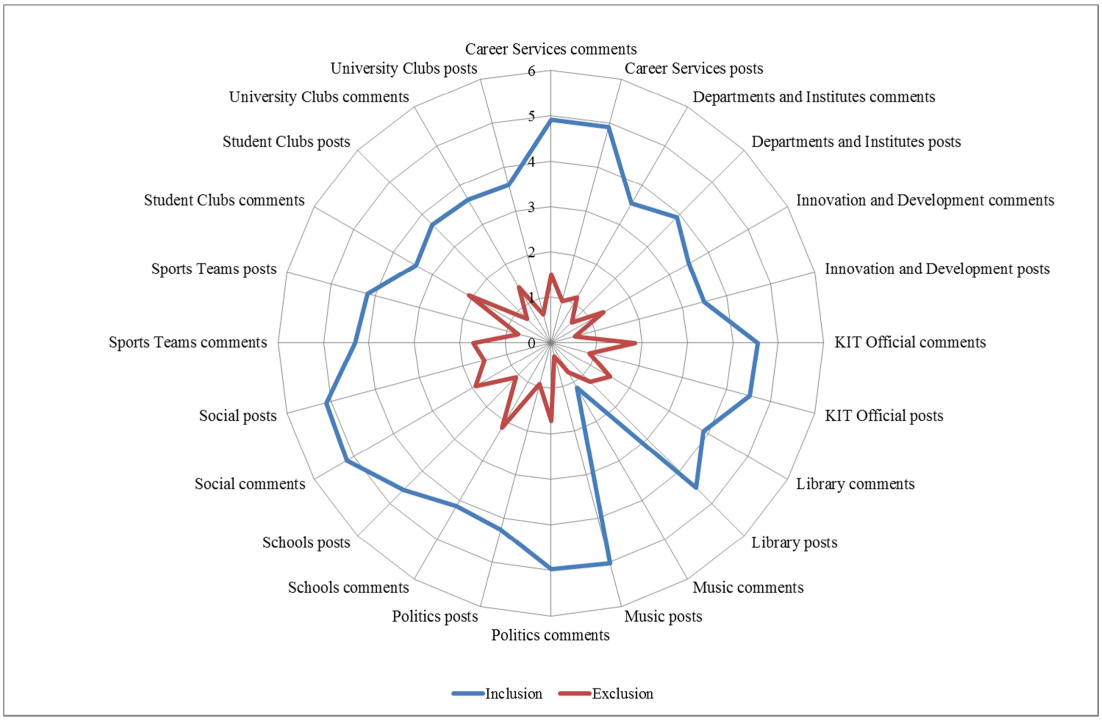


Figure 7.2 Comparative view of inclusive and exclusive speech, posts and comments

Whereas posts are more inclusive than comments, comments are more social than posts (Figure 7.3). In almost all cases, comments spike for social aspects of discourse and posts dip. An exception is the Social posts, where the posts show higher usage of social discourse than the comments. The usage of "Friends" is almost non-existent in this dataset, likely due to the public (as opposed to personal) nature of the KIT Facebook network.

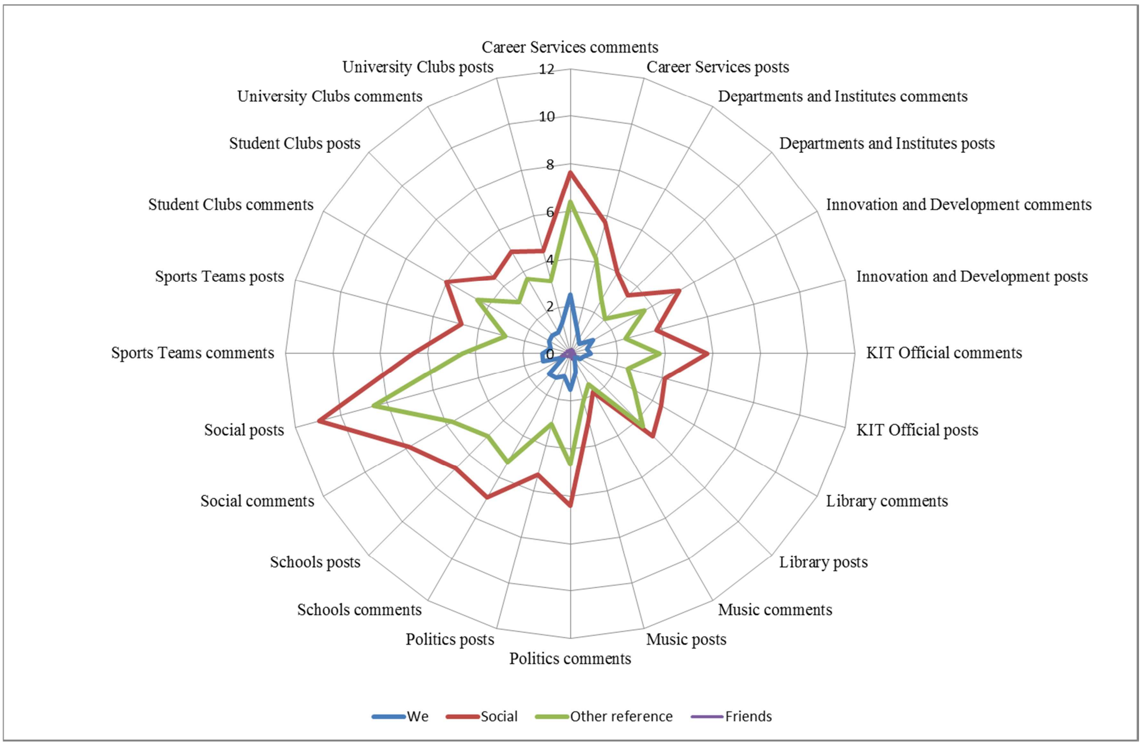


Figure 7.3 Comparative view of social speech, posts and comments

Closely related are the concepts of social belongingness and social status. A strongly hierarchical community will display high levels of status differences, and would likely express low levels of belongingness. As discussed in Section 3.2.3, communal belongingness has been defined as high usage of the categories We, Social, and Inclusion (Figure 7.4). It can be seen that with the exception of the Music comments, both Social and Inclusion are relatively high across the community. Here it should be remembered that the Music pages tended to represent themselves neurotically. First person plural occurs less frequently, meaning that it cannot be taken for granted that the community is a fully cohesive one.

The categories Social Process and Others display remarkable similarities. This is likely due to the similarities of the subjects in the LIWC dictionaries. It is however encouraging seeing that these otherwise similar categories retain their distributions across the posts and comments, indicating consistency in the data.

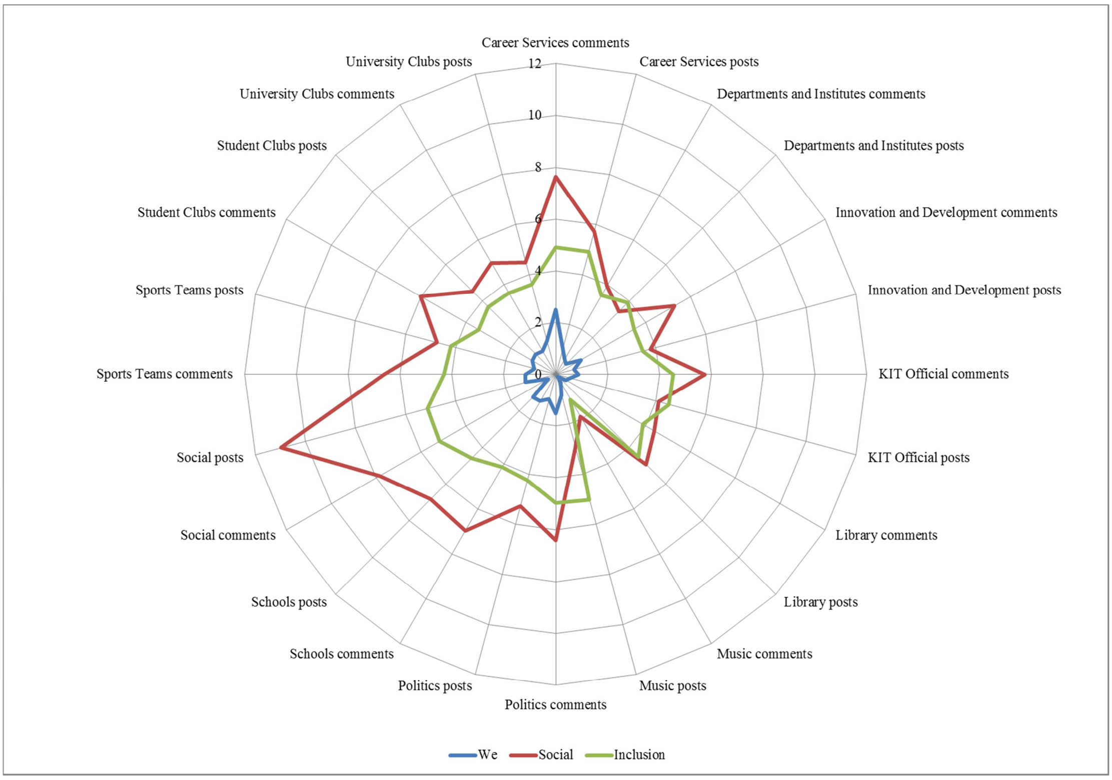


Figure 7.4 Comparative view of communal belongingness, posts and comments

Social status paints a more direct picture (Figure 7.5). Social status is estimated by comparing the frequency of references to others to the frequency of references to self and tentative language. Here it is easy to see that other references occur with a frequency between 2 and 3 times higher than references to self. As tentative language is also low, it can be stated that the KIT Facebook network does not function as a strong hierarchy.

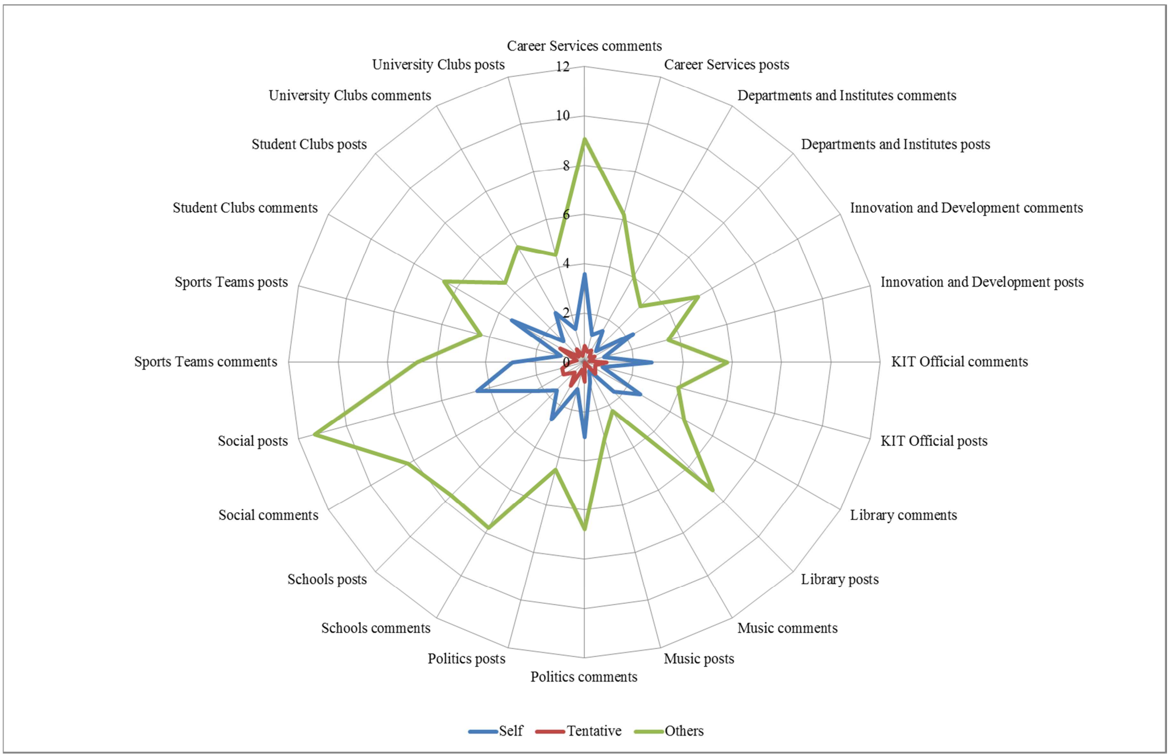


Figure 7.5 Comparative view of social status, posts and comments

The network is also present-focused, which can be understood as a facet of verbal immediacy (Figure 7.6) (Pennebaker, Mehl, and Niederhoffer 2003). This indicates that the discourse on Facebook could tend towards informality. Unfortunately, due to the limitations of the existing dictionary, it is not possible to compare that assumption to the use of formal versus informal person usage (i.e., using the German ‘Sie’ or ‘Du’). Informality is then estimated by following the findings of (Pennebaker and King 1999), who suggest that elevated use of first person singular, present tense verbs, short words, discrepancy words, and the non-use of articles is a marker of verbal immediacy. Verbal immediacy can be understood as a linguistic marker of familiarity (Bazarova et al. 2012). From this metric it is seen that the Social posts and comments are quite informal as well as Student Club comments and posts by the Library (Table 7.4). It is important to note that Library posts are also suspected of engaging in self-representation, and this result for that page group therefore should be read with caution. However the scores hover at or below 0, indicating that while the posts are present-focused, this is unlikely to solely rely on the informality of the discussions.

Table 7.4. Post-comment groups sorted by verbal immediacy metric

Category	Administration or Student-run	Immediacy
Politics posts	Student	-5.93
Sports Teams posts	Student	-5.23
Music posts	Administration	-4.58
KIT Official posts	Administration	-3.75
Career Services posts	Administration	-3.47
Student Clubs posts	Student	-3.43
Schools posts	Administration	-3.27
Departments and Institutes posts	Administration	-2.93
Innovation and Development posts	Administration	-2.29
University Clubs posts	Student	-1.95
Politics comments	Student	-1.7
Career Services comments	Student	-1
Music comments	Administration	-0.76
Sports Teams comments	Student	-0.47
Departments and Institutes comments	Administration	-0.25
Schools comments	Administration	-0.15
KIT Official comments	Administration	-0.02
University Clubs comments	Student	0.29
Library comments	Administration	0.45
Innovation and Development comments	Administration	0.8
Library posts	Administration	1.93
Social posts	Student	2.51
Student Clubs comments	Student	2.82
Social comments	Student	2.87

Similarly to Chapter 5, the lack of Future tense is surprising (Figure 7.6). One could assume that students and the administration use Facebook to alert others about upcoming events (e.g., sporting or musical events, parties) and opportunities (e.g., scholarship deadlines from the Schools and Departments), but this appears to be untrue. The only case where Future exceeds Past is from Music Posts, but even here Present use exceeds Future use.

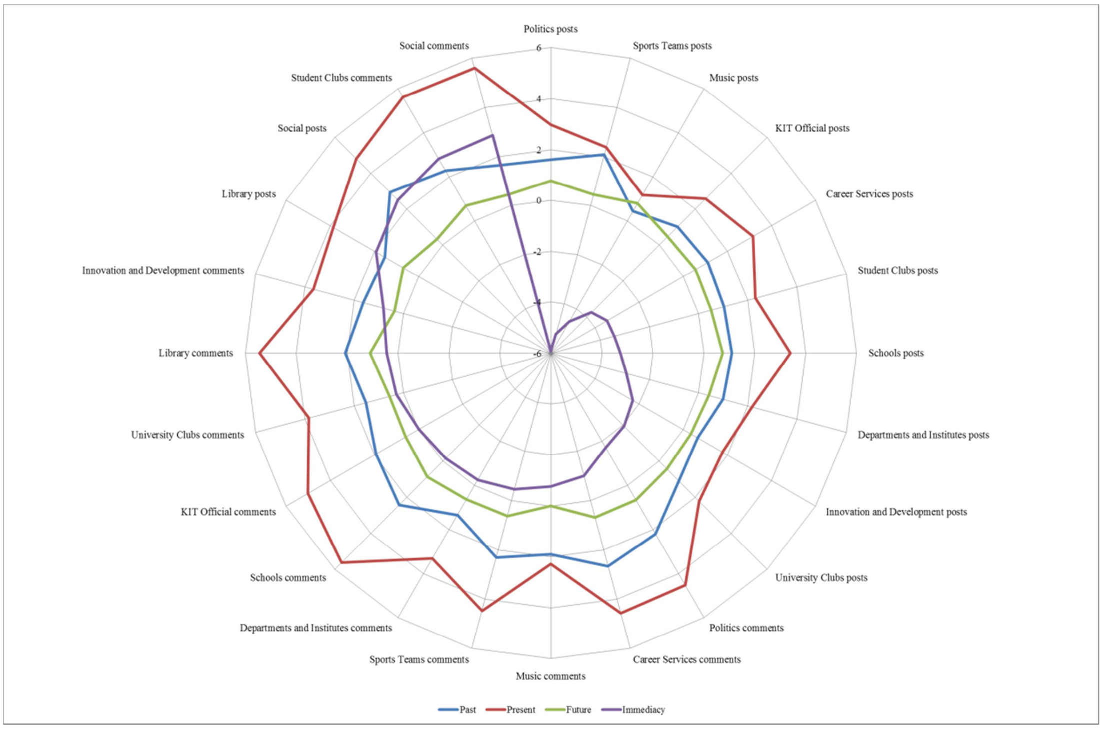


Figure 7.6 Comparative view of the use of tense in speech, posts and comments sorted by the factor immediacy

When considering professional discussions (Figure 7.7), not only the Career Service pages have spikes in career related topics (a frequency of 6.28%), but also the Schools of the university as well (6.61%). Quite unexpectedly, the politically inclined groups have equal references to career-related aspects to the Schools, which is even higher than the Career Services pages (6.37%) (though this is statistically insignificant). References to Jobs spike in posts, indicating that the pages are attempting to sponsor career opportunities. Several notable patterns appear in the comments: for the Sports comments, Achievement and School are equal. In the Political commentary, Job and Achievement are equal. And, the commentary on the Library pages reference School, Jobs, and Achievement with equal frequency: which is to say, infrequently in comparison to the rest of the post and comment groups.

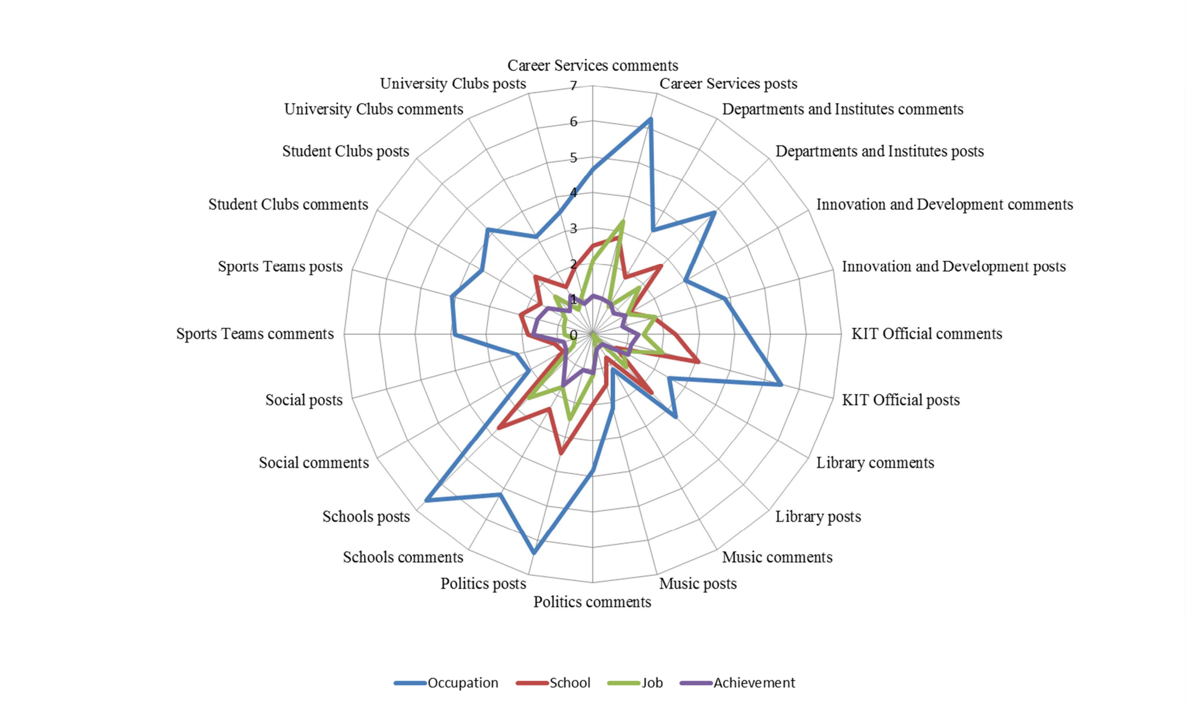


Figure 7.7 Comparative view of professional speech, posts and comments

Comparing posts and comments reveals interesting differences in the discourse baseline. Positive Emotion (mean= 2.56, SD 2.13) is used more frequently than Negative Emotion (mean = 0.577, SD 0.667) in line with the findings of the previous chapters and (Pennebaker, Mehl, and Niederhoffer 2003). Results of an Independent Sample Mann-Whitney U test show highly significant differences in the use of Positive Emotion (U = 6,740, z = -4.520, p = .0005) and Negative Emotion (U = 7,530, z = -3.381, p = .0005), using an asymptotic sampling distribution for U. Mann-Whitney U is the non-parametric estimation of a One-Way ANOVA. Figure 7.8 illustrates the mean differences in usage; comments show a higher frequency of more positive and negative emotional discourse. When these emotions are employed, they tend to be employed in comments.

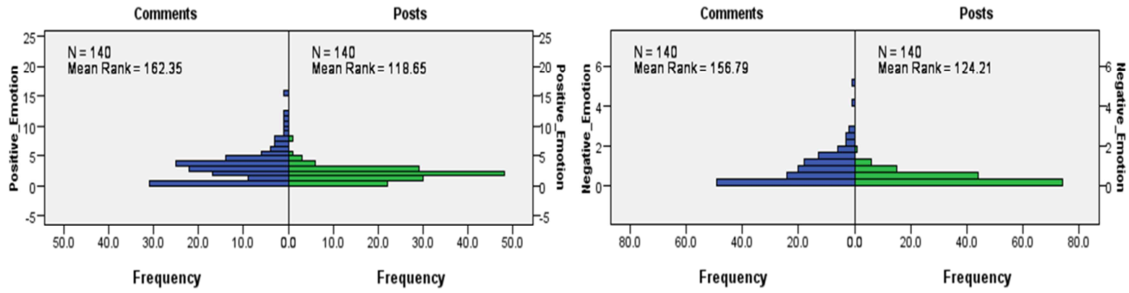


Figure 7.8 Results of a Mann-Whitney U test comparing usage of Positive and Negative Emotion

Again, Net Affect is calculated by subtracting negative sentiment categories from positive sentiment categories (see Section 5.5.3 for a description of this). Compared to Chapter 5's negative Net Affect across Facebook discourse, the KIT network is mesokurtic with a positive skew (Figure 9.8a) and a reversed sigmoid distribution (Figure 7.9b), hovering at zero but with a long positive tail.

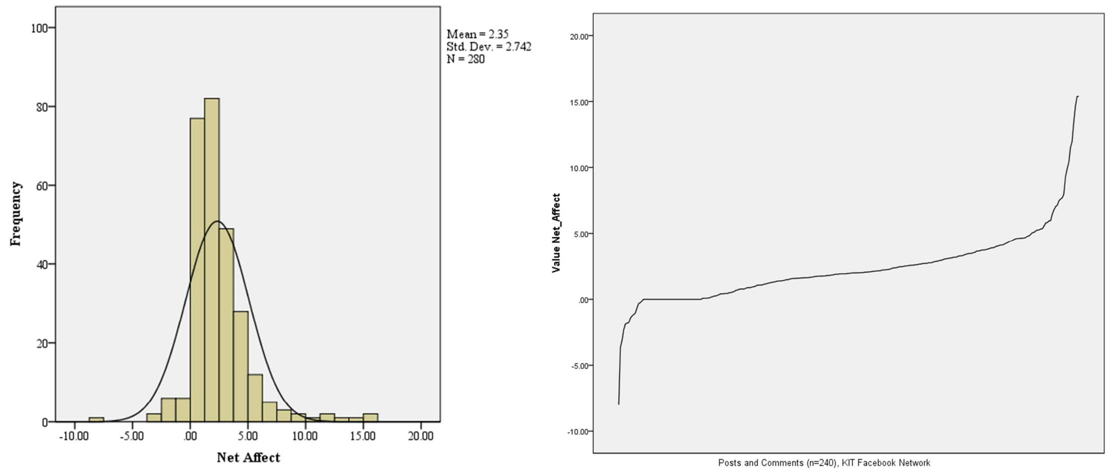


Figure 7.9 Net Affect, displaying skewedness and (a) Kurtosis and (b) Distribution

That KIT's Net Affect tends to hover around zero signifies few pages employing extreme emotion. The absolute range is -8.0 from the OSKar- Optics Students Karlsruhe e.V comments to a positive 15.38 from the comments of the Institute of Regional Science. Comments tend to make up both ends of the tails, and posts are grouped in the middle of the distribution (the zero range). This supports the results of the Mann-Whitney U tests that comments are move emotive than posts.

Visualizing Net Affect as a relationship graph has telling features. Figure 7.10 is the relationship graph of KIT's expressed well-being, showing the weightiest edges. The posts of the KIT main page's posts maintain a fairly central position that is interestingly neither connected to posts by the Schools of KIT, nor its comments. Density in relationship to KIT posts is rather by similar well-being expression profiles with Career, Politics, Innovation and Development, and University Clubs. The KIT main page comments are situated near comments on Politics, Schools, University Clubs and posts on Sports groups. A small cluster between the comments of Career, Student Clubs, Innovation and Development, and Sports is also visable. This is a likely indication that the commenters of these pages have overlapping interlocutors. Interesting is the lack of connectivity with the Social comments and Music comments. While Music comments shares a similar profile with Music posts, the Social comments are completely isolated from the network. A visual inspection of the data reveals that while Social comments do not have the most extreme distances, the distances between these comments and other is consistently higher than all other pairings.

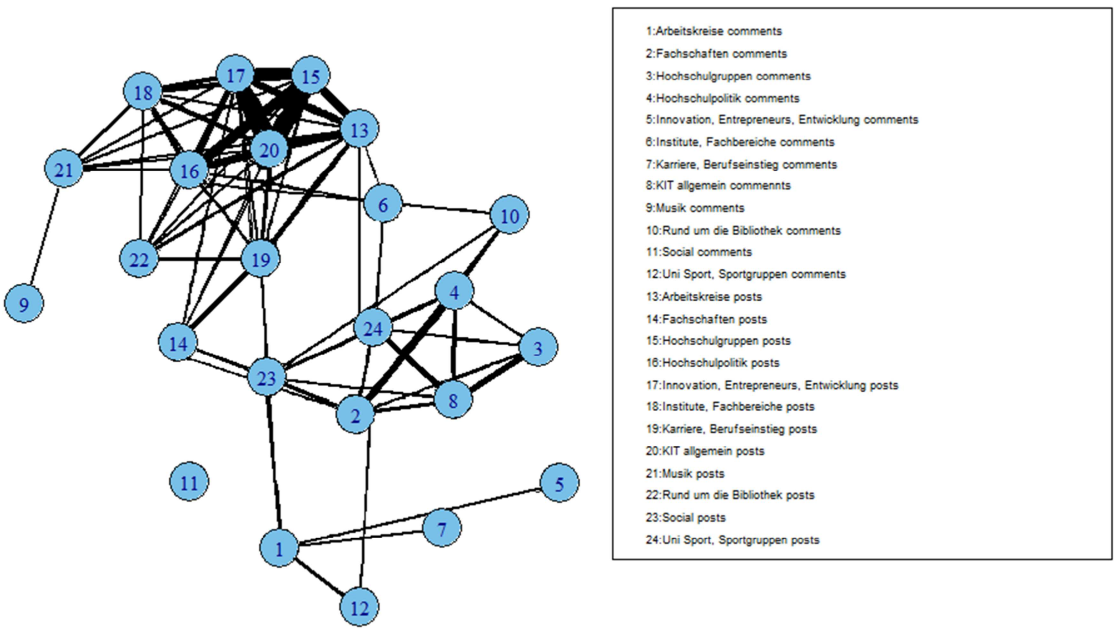


Figure 7.10 KIT's well-being relationship graph

Agreement level is also an interesting characteristic of university discourse. There is a highly significant different in the way that Assent (U = 6,691, z = -4.688, p = .005) and Negation (U = 7,366, z = -3.611, p = .005) are used according to an asymptotic sampling distribution Mann-Whitney U Test (Figure 7.11).

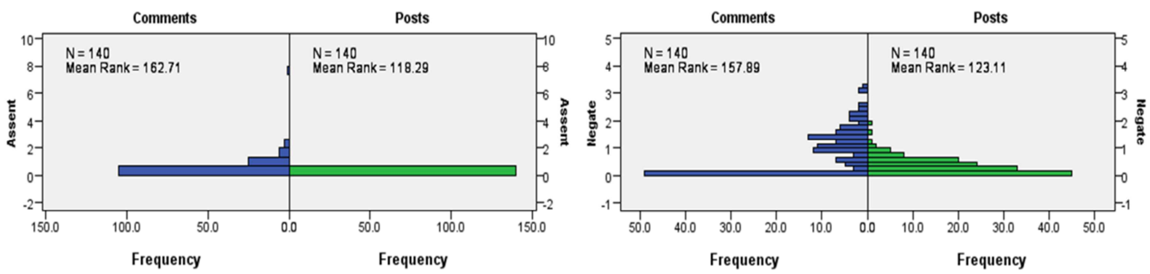


Figure 7.11 Results of a Mann-Whitney U test comparing usage of Assent and Negation

Comments are reactive to posts. The frequency of Negations is highest in comments; Assent is likewise more frequently expressed in comments. This finding is reflective of comments being likely to discuss the topics mentioned in the preceding post. When this is considered alongside with the tendency of comments to use more cognitively expressive and emotive discourse in their responses (Figure 7.12), it can be understood that although this tendency should be expected in most communities, the size of this gap indicates that the university's constituents visit the pages to seek and engage in lively discussions. Comments display significantly higher cognitive complexity than posts (U = 5,831.5, z = -5.861, p = .005).

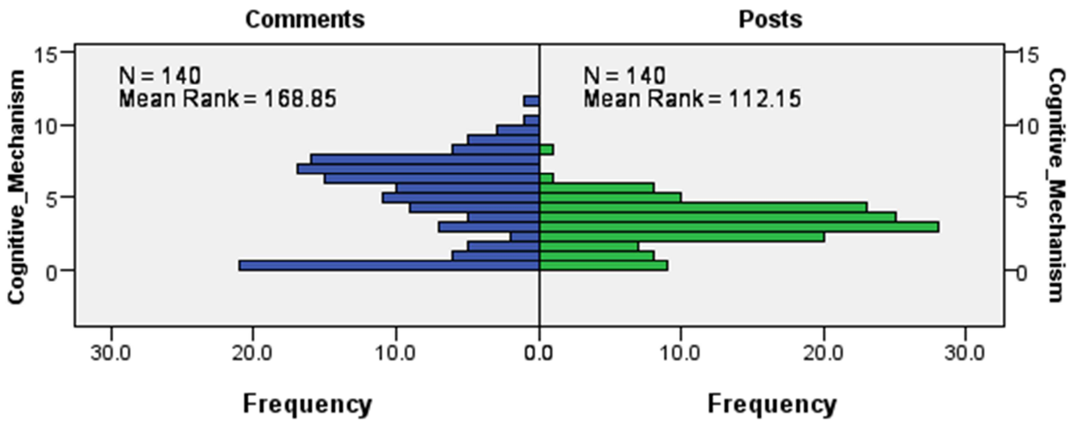


Figure 7.12 Results of a Mann-Whitney U test comparing cognitive complexity

Linguistic Accommodation

Linguistic accommodation signals high degrees of engagement between and amongst discourse participants (Niederhoffer and Pennebaker 2002). The indication that comments are reactive to posts existing in the Facebook communication is a positive finding, suggesting that community members are quite responsive and engaged with one another. In Chapter 5 it was established that linguistic accommodation did not occur due to the rapidly changing discussion partners in a given Facebook exchange. However, comments imitate a one-turn mutual chat interaction between posters and commenters in the KIT use case. Therefore the next research aspect to be covered is the hypothesis of linguistic accommodation (Danescu-Niculescu-Mizil, Gamon, and Dumais 2011; Niederhoffer and Pennebaker 2002).

To investigate the existence of linguistic accommodation, first an estimate of dissimilarity per page group is taken using a Euclidean distance analysis (Table 7.5). Comments have an average dissimilarity of 5.7 and posts have an average dissimilarity of 6.62. Post-comment combinations have an average dissimilarly of 8.88. The average dissimilarity between page groups is 7.37, with a SD of 3.86. Page groups with a dissimilarity score below 3.51 (the SD subtracted from the mean) show high linguistic accommodation, as low dissimilarity scores as tantamount to higher similarly within the dataset. Fitting this description are 13 pairs:

Table 7.5. Linguistic Accommodation, estimated via Euclidean distance

Sports Teams posts	Social posts	Library posts	Music posts	KIT Official posts	Career Services posts	Departments and Institutes posts	Innovation and Development posts	Politics posts	University Clubs posts	Schools posts	Student Clubs posts	Sports Teams comments	Social comments	Library comments	Music comments	KIT Official comments	Career Services comments	Departments and Institutes comments	Innovation and Development comments	Politics comments	University Clubs comments	Schools comments	Student Clubs comments	
8.84	10.46	7.19	13.75	9.73	8.38	10.89	11.14	9.85	9.70	7.96	9.12	3.28	5.06	4.39	14.24	3.48	5.00	7.98	5.55	4.02	4.96	4.40	0.00	Student Clubs c
9.28	10.20	7.80	14.70	9.52	7.31	11.26	11.96	8.32	10.51	6.10	9.50	5.05	6.43	5.75	15.99	3.85	4.85	9.18	7.97	3.62	7.07	0.00		Schools c
6.04	12.49	4.89	10.31	6.25	6.13	6.72	6.60	7.84	5.23	6.97	5.06	4.80	6.56	4.43	9.95	4.26	6.63	3.64	1.88	5.37	0.00			University Clubs c
8.03	10.67	7.00	13.16	8.41	6.78	9.93	10.56	7.86	9.07	6.52	8.09	4.36	5.65	4.41	14.15	2.35	4.18	7.53	6.40	0.00				Politics c
6.81	12.91	5.74	10.92	7.10	6.90	7.33	6.89	8.76	5.73	7.76	5.81	5.37	6.97	5.59	10.18	5.35	6.93	4.48	0.00					Innovation and Development c
5.91	14.96	5.26	8.86	4.54	6.15	3.98	4.06	7.04	3.23	7.71	3.33	7.93	9.25	6.16	7.55	6.43	9.20	0.00						Departments and Insititutes c
9.80	9.66	8.23	14.85	10.13	7.75	11.45	11.70	9.54	10.20	7.01	9.58	5.15	5.79	7.26	15.55	5.06	0.00							Career Services c
7.11	11.58	6.12	12.40	7.25	5.94	8.76	9.50	7.17	8.13	5.96	7.14	3.80	6.01	3.63	13.26	0.00								KIT Official c
10.46	20.10	10.59	9.57	9.64	12.18	7.62	6.71	12.57	7.32	14.02	8.38	14.09	14.89	12.22	0.00									Music c
7.23	12.32	6.41	11.86	7.95	7.60	8.97	9.37	8.24	8.16	8.00	7.43	4.98	6.63	0.00										Library c
10.90	8.51	7.81	14.91	11.29	9.55	12.24	12.11	11.74	10.65	9.24	10.47	5.85	0.00											Social c
7.02	10.49	7.10	12.68	8.86	7.46	10.27	10.51	8.88	9.03	6.95	8.30	0.00												Sports Teams c
4.29	15.00	5.37	7.61	2.41	4.45	2.55	3.02	5.09	2.02	6.16	0.00													Student Clubs p
6.77	12.03	6.27	11.79	5.72	2.88	7.73	8.54	4.59	7.39	0.00														Schools p
5.57	15.11	5.52	7.64	3.97	5.76	2.98	2.14	6.86	0.00															University Clubs p
5.49	15.28	7.60	10.33	3.90	4.05	5.94	7.42	0.00																Politics p
6.37	16.75	6.78	7.61	4.23	6.61	2.34	0.00																	Innovation and Development p
5.72	17.08	6.74	7.69	2.52	5.68	0.00																		Departments and Institutes p
5.86	13.47	5.83	10.54	3.75	0.00																			Career Services p
4.52	15.86	6.09	8.28	0.00																				KIT Official p
8.11	18.64	9.76	0.00																					Music p
7.00	12.07	0.00																						Librarby p
15.14	0.00																							Social p
0.00																								Sports Teams p

- Student Clubs comments – KIT Official comments
- Sports Teams comments – Student Clubs comments
- KIT Official comments – Politics comments
- Departments & Institutes Comments – Student Clubs posts
- Departments & Institutes Comments – University Clubs posts
- Student Clubs posts – University Clubs posts
- Student Clubs posts – Departments and Institutes posts
- Student Clubs posts – KIT Official posts
- Schools posts – Career Service posts
- University Clubs posts – Departments and Institutes posts
- University Clubs posts – Innovation and Development posts
- Departments and Institutes posts – Innovation and Development posts
- Departments and Institutes posts – KIT Official posts

Notable is that there are no post-comment page pairings. This indicates that while it is likely that groups of commenters can be identified, and which pages have similar posts, it is not possible to identify linguistic accommodation in this dataset. This is reasonably due to the same factors as seen in Chapter 5; discussion partners change too rapidly (or anonymously) for linguistic accommodation to take root.

Deceptive Language

Another factor to consider for this network is the propensity to engage in deception or deceptive conversation patterns. The analysis of deceptive statements is based on the findings that liars express less first person singular (‘I’) and more ‘negative emotion’ due to feelings of guilt evoked by the act of lying, and depict less cognitive complexity as capacity is needed to establish a convincing story, reflected by fewer ‘exclusion’ and more simple ‘motion’ words (Newman et al. 2003; Ott et al. 2011). While there is little immediate incentive for outright lies in such a network, there can be various drivers for deceptive actions. Especially lies of administrators are more vulnerable to be detected as the pages’ official actions tend to be publicly observed with higher interest compared to individual comments. Thus, a single witness of contradictory information could reveal deception to the whole community and page administrators are expected to be aware of this fact. Some examples of reasonable deceptive practices could be page administrators seeking positive feedback, publicity or attention could try to support these achievements by drastically exaggerating or even ‘making up’ interesting stories. Individual commenters could aim at receiving the community’s recognition and based their deceptive actions off of this. Whilst page administrators often form teams and lies may require collective consent, individual page commenters in the KIT community enjoy high anonymity, facilitating untruthful statements.

Two possible methods exist for the assessment of deception in Facebook discourse: direct score comparison (i.e., as done to assess self-representation) or a summed approach (i.e., as seen in the calculation of Net Affect). Given the simplicity represented by single scores, the additive approach is chosen. Thereby two deception metrics are established: The sum of ‘I’ and ‘exclusion’, as well as ‘negative emotion’ and ‘motion’ LIWC scores. ‘Negative Emotion_Motion’ is subtracted from ‘I_Exclusion’ to reach a single score. The baseline values for the two scores are measured separately for all posts and comments as the chosen categories show sizable gaps between entry types, most likely resulting from differing basic characteristics (for an overview of LIWC scoring see Section 3.2.3). The second SD is again chosen to establish baseline differences as one to one comparisons between the two scores would lead to identifying almost every post as deceptive due to standard smaller values for ‘I’ and Exclusion, thus reflecting a logical mistake by ignoring the purposes of each entry type (e.g. a common purpose of posts is to evoke discussion and of comments to give personal opinions). If an individual post or comment demonstrates both, a near absence of frequency of ‘I_Exclusion’ use and an exaggerated frequency of ‘Negative Emotion_Motion’ compared to the according baseline-scores of the database, it is tagged as highly suspicious. To reduce variance only entries with length of 35 words (the average sentence in German) or more are considered. This restriction further respects that lie detection depends on a reasonable amount of linguistic information.

Two granularities are investigated. First pages are split based on the type of page administration: university administrator led pages, or student led pages (Figure 7.13). The administration-student management granularity is well suited to deliver insights on deceptive post-comment comparison. Despite of the above mentioned barriers for page managers to share exaggerated or wrong information, the established deception rate almost doubles from comments to posts, reflecting a rather unexpected finding. One explanation would be people accepting and expecting certain levels of overstatements in posts on Facebook pages. This discrepancy is left for future work.

Focusing on relative increases due to the differences in dataset sizes is also necessary. Officially administrated pages show highly suspicious posts for 478 out of 4586 possibilities, equaling 10.4%. Deception marginally increases (11.4% increase) when students are authorized to manage pages resulting in a total deception proportion of 11.6%. This finding holds true for commenters as well: Commenters on student-run pages present an 18.8% higher occurrence of possibly deceptive comments (6% from 5.1% on employee-administrated pages). Seemingly, student administrators respect the responsible position slightly less honestly than administration employees of the KIT. Additionally, administration-led pages influence commenters’ tendency to write possibly untruthful statements. The analysis of Chapter 7.2.2 has established that student-run pages evoke a less formal environment for visitors. This aspect may reduce visitors’ inhibition to lie on student-run pages.

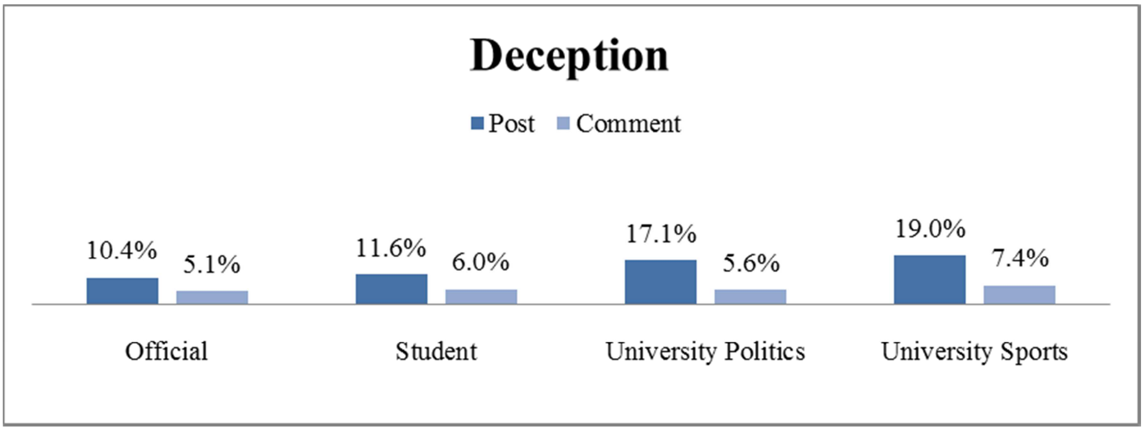


Figure 7.13 Frequency analysis of deceptive-type comments and posts

The second granularity investigated is the page groups as explained in Section 7.2.2. Each supra-group (KIT (official presence), Library, Schools, Departments and Institutes, Student Clubs, University Clubs, Sports Teams, Innovation and Development, Politics, Carrier, Music, and Social) is individually assessed across posts and comments. Generally this analysis did not show high levels of deceptive aspects. 20 out of 24 possible post-comment groups had only marginal posts or comments which could be considered deceptive. Perhaps unsurprisingly Politics-related posts (17.1%) and comments (5.6%) contain above average deception rates (Figure 7.12). The other post-comment group presenting exceptionally high proportions of suspicious content for posts (19%) and comments (7.4%) is the Sports Teams pair (Figure 7.12). These pages mainly feature game reports of diverse university teams. Here it is reasonable to assume that hard lies about results would not appear, but rather exaggerating positive performance in case of wins and underplayed reasons for defeat when a match is lost might be prevalent.

7.2.3 Temporal Representations

Considering fast-paced online communities there is an interest in knowing if, and which, events have notable effects on the way the community interacts, and if there are sentiment changes over longer time periods. One way to identify events of impact is to visually inspect spikes and dips as they are related to the semester intervals. With the semester intervals acting as a baseline, obvious highs and lows in communal sentiment are more easily identifiable. Temporal representations are segments of the datasets parsed for different, small time periods within the larger semester timeframe. The following analysis address the benchmarks of the semester, highlight two events that are especially noticeable in emotive spikes from the data, and names other events which were expected to correspond with increased latent emotion but had no visible or statistically significant impact on the KIT community discourse.

In order to create a comparative baseline, LIWC scores of all data (posts and comments) before the start of the event and after its completion have been aggregated to a single number, weighted by total word counts. Considering time-local tendencies, the three equal time intervals of one month before and after, and the month during games are analyzed. All measures in the coming analyses do not show the actual LIWC scores, but relative increase and decrease to the baseline.

Semester Intervals

The KIT community is highly cyclic, as noted in Section 7.2.1. Figure 7.14 displays an average of the academic year considering the timespan 2011-2014. There it can be seen that the bulk of discussions occur inside of the semester, with the Winter Semester having slightly more chatter than the Summer Semester. This pattern is flipped for the holiday seasons, which Summer Holidays having a slight boost in activity compared to the Winter Holidays. That remains constant when comparing the exam weeks to the holidays – Winter Holidays have less Facebook interaction than the Winter Exams, and Summer Holidays have more interaction than the Summer Exams.

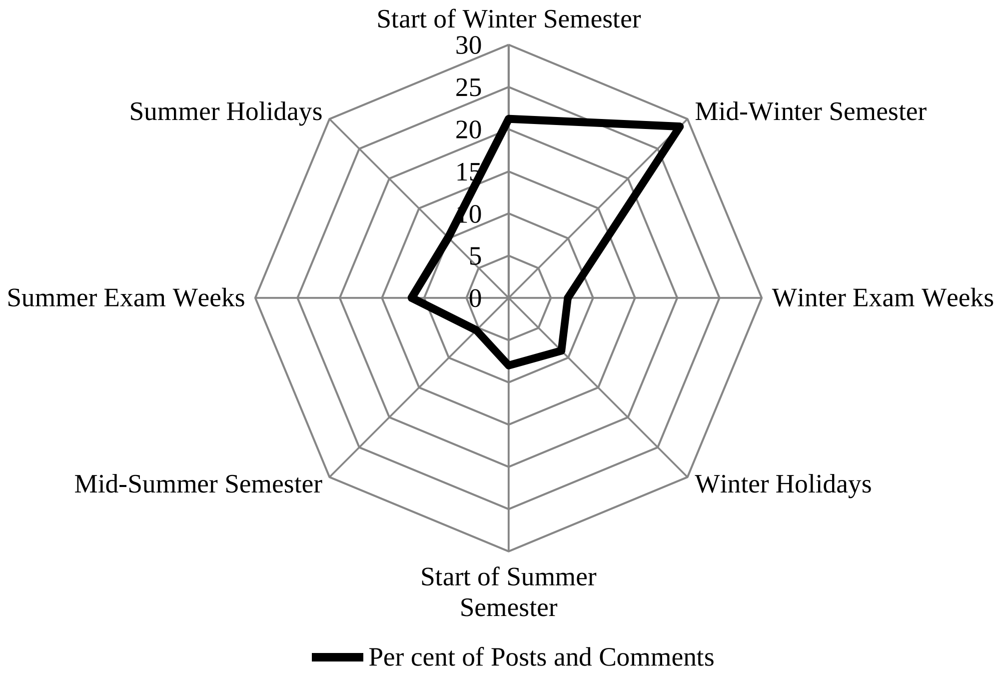


Figure 7.14 Frequency of KIT posts and comments throughout the academic years 2011-2014

Discernable patterns are found in the expression frequency of positive and negative emotions that coincide with the semester calendar (Figure 7.15). Likely due to the influence of

Christmas and New Year's, Positive Feelings are highest during the Winter Holidays. Anxiety is lowest during the semester holidays and highest during the summer term. Anger and Negative Emotion are most common inside of the winter semester; Sadness and Optimism are most common inside of the summer semester.

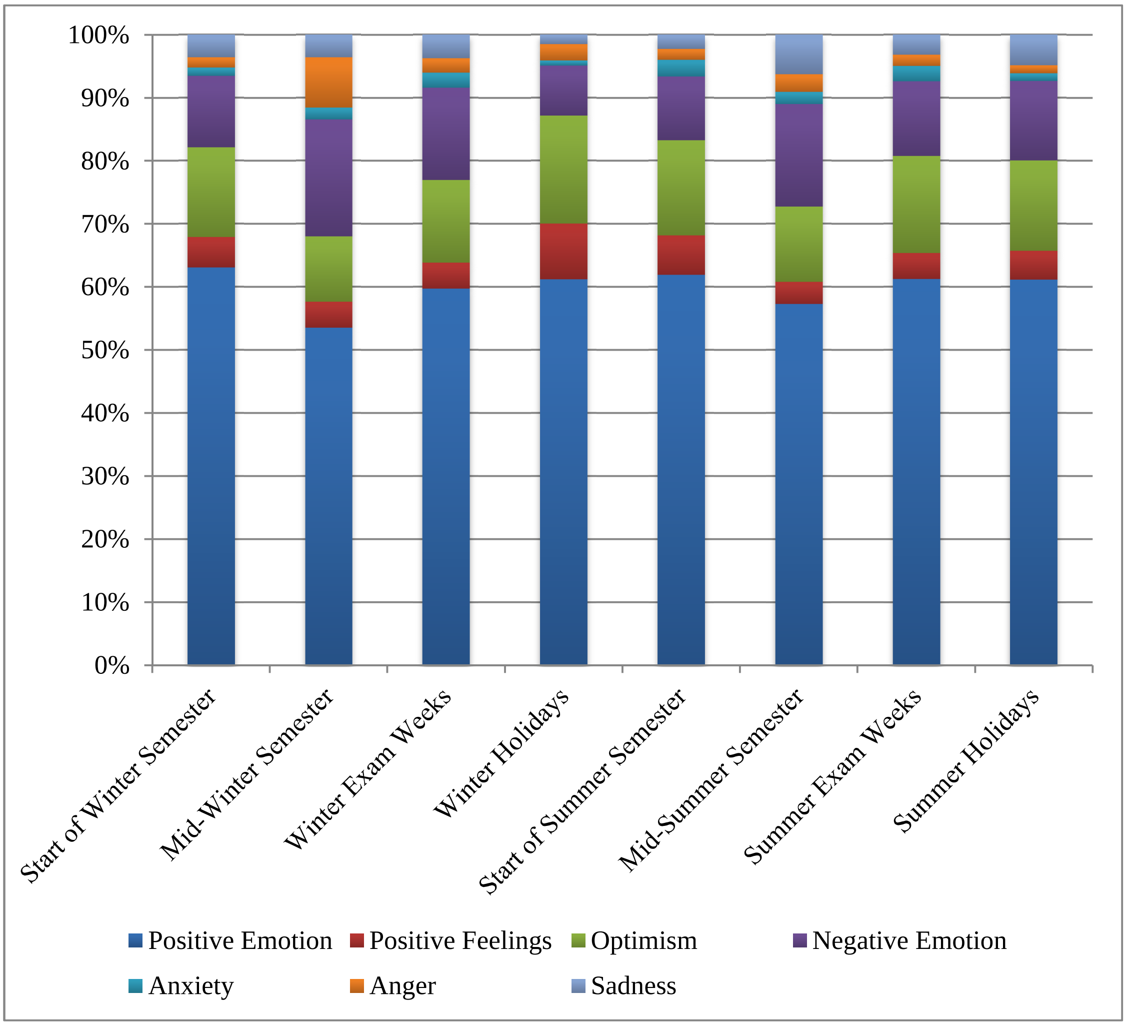


Figure 7.15 Frequency of KIT posts and comments throughout the academic years 2011-2014

Additionally, results show peaks during the semesters for the categories Cognitive Mechanism and Social Processes (Figure 7.16), and decrease during holidays and exams. This could be influenced by the logic assumption of students interacting most when lectures are in full process and no additional stress is put on them. That Cognitive Mechanisms are lower inside of the semester than during exams is likely due to decreased network engagement by students.

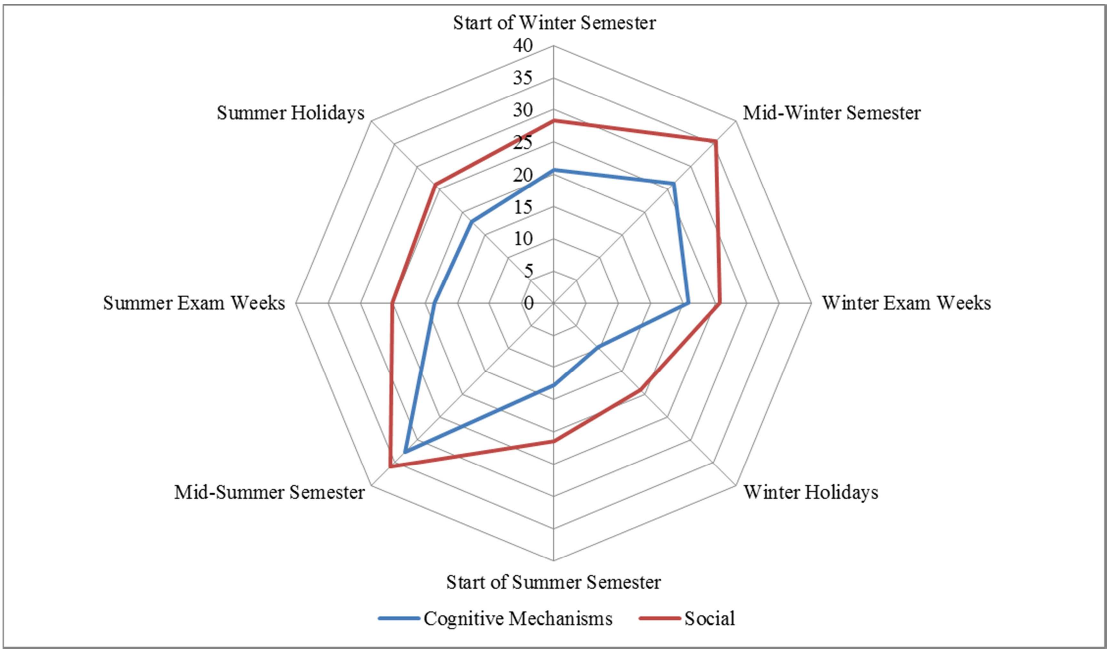


Figure 7.16 Frequency of cognitively oriented discourse and social discourse throughout the academic calendar, 2011-2014

Germany's Excellence Initiative II

BeWell@KIT established a critical disappointment for students and employees as the denial of the Elite Status on 15 June 2012.[41] The loss acted as a shockwave across the network and was the most common discussion topic the days after the loss, as it was expected to damage the university's prestige and also included the end of the additional 'Excellence Money,' a governmental financial support of 15 to 20 million euros yearly.[42] Since the first round of funding in 2006 the KIT proudly presented its Elite status, a national governmental award for scientific research of the highest quality. Students feared decreasing employment opportunities in the highly competitive academic working environment. At the same time, financial consequences threatened the continuing of research projects and existence of administration jobs. Hence, the denial impacted students, researchers, and administration employees likewise.

First a strong rise in the Facebook community's overall activity can be seen after publication of the judges' Excellence decision. Whilst the week before the announcement counts 7,425 words, this amount increases by one third to 11,070 words during the consecutive week and 15,072 (almost an additional 25%) two weeks after the event. The two weeks representing the event and after the event comprise 1.3% of the four years of corpus' words. The categories reflecting cognitive complexity (Articles, Exclusion, Causation) show a positive trend in the following week of the Excellence loss compared to the overall score before (Table 7.6).

[41] http://www.kit.edu/kit/english/5963.php. Last Accessed: 3 January 2015.
[42] http://www.deutschlandfunk.de/nicht?mehr?exzellent.680.de.html?dram:article_id=240282. Last Accessed: 3 January 2015.

Putting this together with the significantly higher scores of Past and Future (measuring verb tense frequency), and the topic categories Money, Occupation, Job and School is an indication of intense discussion on the reasons and future impacts of the Elite denial.

Table 7.6. Score development for comparison between 1) all data before June 15th 2012, 2) the following first week after the event and 3) the following three weeks after the event where green shows increases and red shows decreases

	Before Loss	1 Week After	3 Weeks After
Articles	6.68	8.24	7.64
Exclusion	0.86	1.04	1.04
Causation	0.63	0.88	0.72
Past	1.31	1.85	1.71
Future	0.56	0.78	0.71
Money	0.72	0.89	0.68
Occupation	5.49	6.07	5.83
Job	1.89	2.06	2.04
School	2.87	3.37	3.19

It is a promising and intuitive finding that the first week shows the most distinct peaks for all cases. Still, a wider timeframe post-event produces the same tendencies for all LIWC categories but Money (Figure.7.17). The additional three-week timeslot enables observation whether detected peaks presume or ebb away quickly. Sentiment dimensions seem to differ on the durability characteristic, as some scores almost plateau over three weeks (Exclusion, Past, Future, Job) and others drop back to the benchmark rapidly (Money).

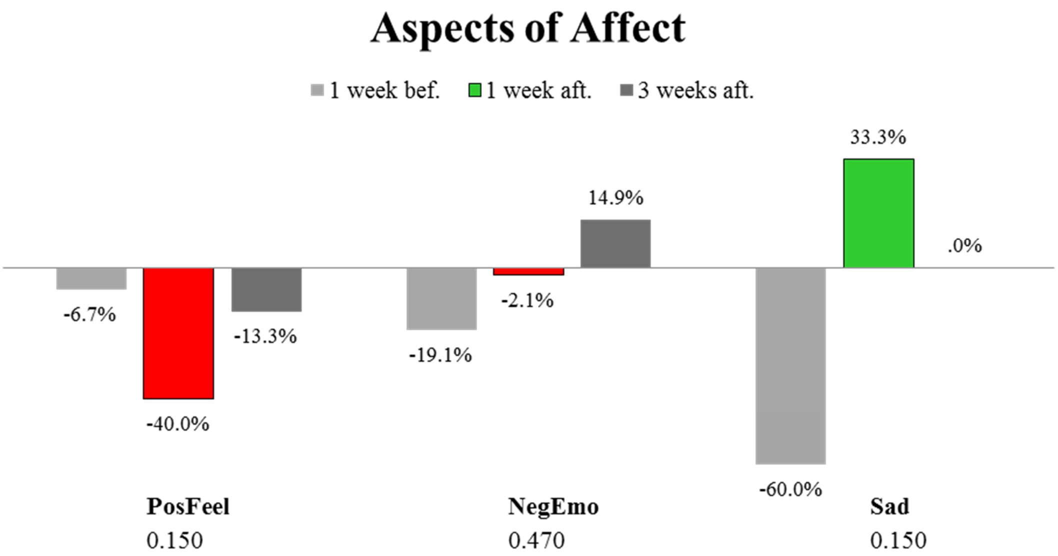


Figure 7.17 Affective changes in discourse relating to the KIT Elite loss. All measures show relative changes, not absolute LIWC scores. The colored bars in the middle reflect the crucial short-term results, while bars to the left (1 week before) and right (3 weeks after) improve interpretation by considering temporal deviations from the baseline and resilience of effects.

More than impacting professional and practical concerns, the loss of the Excellence status had a major influence on the KIT's digital expressions of well-being. Increased frequencies of the categories Negative Emotion and Sad hint at a frustrating occurrence around June, 15th. Positive Feeling depicts a decrease (-35.7%) directly after announcement of the denial. It is interesting to observe that after the first distinct drop, zooming out to the following three weeks, the category shows a slight upswing indicating communal resilience while reminding us how delicate results based on latent emotional states are (Figure 7.18). The LIWC category Social increases slightly after the incident, and quickly increases in the following three weeks. In addition, Inclusion depicts a typical spike as compared to the results in Table 7.6. Inclusive speech then plateaus for the weeks following the event.

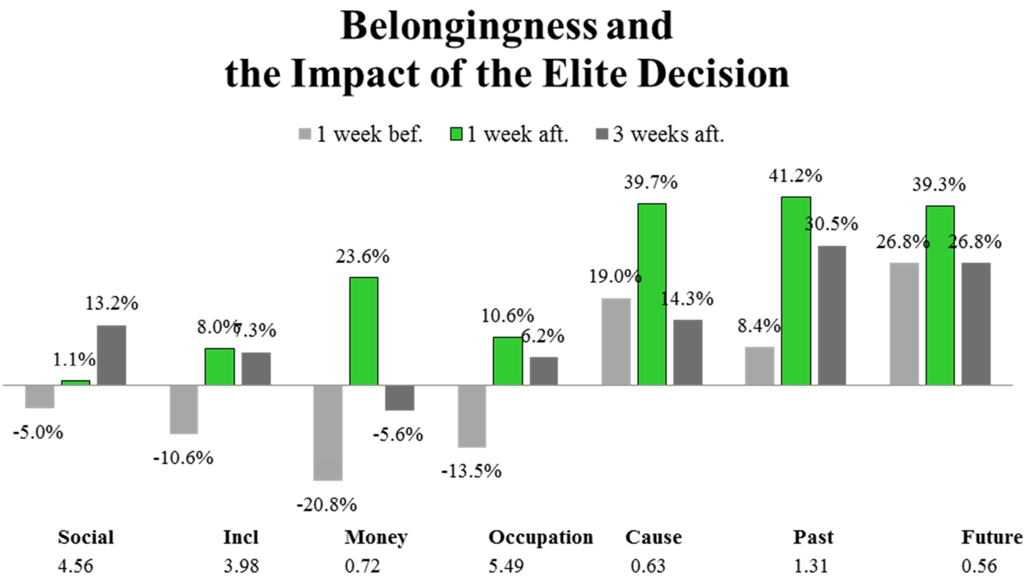


Figure 7.18 Emotive sentiment flow in discourse relating to the KIT Elite loss.

These two categories are strong reflectors of communal belongingness, thus leading to an interesting finding. Because the loss was unexpected it affected almost all community members: the shock was wide-spread and deep. Former research found that tragic collective experiences often promote feelings of belongingness (Pennebaker, Mehl, and Niederhoffer 2003; Kramer 2010; Pennebaker and Lay 2002). This is evident in the KIT dataset, where the loss of the Excellence status acted as a collective crisis according to the Facebook discourse. Encouragingly, the community responded with not only shock and negative feelings, but also resilience and an increase in togetherness, signs of well-being according to the definition of Huppert and So (2011).

Inspection of the 2012 Excellence initiative suggests that campus-wide incidents affect the way the community interacts. Well-being and communal belongingness are affected in the short-run, but the long-term impacts are minimal. This highlights both communal resilience, and how delicate the results are.

World Cup 2014

The 2014 World Cup competition dominated international (traditional) media during the time span 12 June- 13 July 2014, and the World Cup final between Germany and Argentina evoked 280 million interactions by 88 million people on Facebook, which is record for a single sports game.[43] In addition soccer is the most popular national sport in Germany along with most of the world. Germany's 2014 performance and finally becoming world champion for the first time since 1990 resulted in exuberant nation-wide celebration. The final was viewed by 34.65 million people in Germany alone.[44] Therefore, it is not surprising that as an event of interest for both the campus and beyond, and that it registered on the BeWell@KIT sentiment indicators. A single category, Sports, covering 28 sportive expressions, provides evidence that it can be used to detect mega events, with a 42.1% increase during the month of the games.

Excitement and anticipation of games increased frequency of emotive statements as seen by the relative LIWC score rise of 9.8% in Affect (Figure 7.19). This is met by significant changes in the sentiment categories Positive Emotion (+14.0%), Positive Feelings (+46.7%), Negative Emotion (-10.8%) and Anxiety (-30%) the month of the World Cup. Decreasing negative expressions is an especially telling result. Whilst the raise of positive scores could be restricted to posts directly referred to games, the decrease of negative latent emotion indicates an overall sentiment shift to higher community well-being. This is in line with the findings of eminent well-being researchers like Ed Diener, Daniel Kahneman, and their colleagues who find that well-being is not only the presence of positive emotions but the absence of negative emotion (Diener 1984a; Kahneman and Krueger 2006). A conflicting result appears for the LIWC category Sad. Various reasons for the increase could be based on a logical relation to the games. Some reasons could be that the campus is an international environment and also there are many natives rooting for other favorite soccer teams; also the games took place six time zones from Germany, which meant that the schedule conflicted with a daily work-life schedule as well. Reasonably, there is some possibility of sadness because of empathic statements for losing teams in the case of otherwise good game performance.

[43] http://newsroom.fb.com/news/2014/07/world-cup-breaks-facebook-records/. Last Accessed: 20 January 2015.

[44] http://www.presseportal.de/pm/6694/2783889/das-erste-neuer-rekord-34-65-millionen-zuschauer-sahen-fu-ball-wm-finale-deutschland-argentinien. Last Accessed: 20 January 2015.

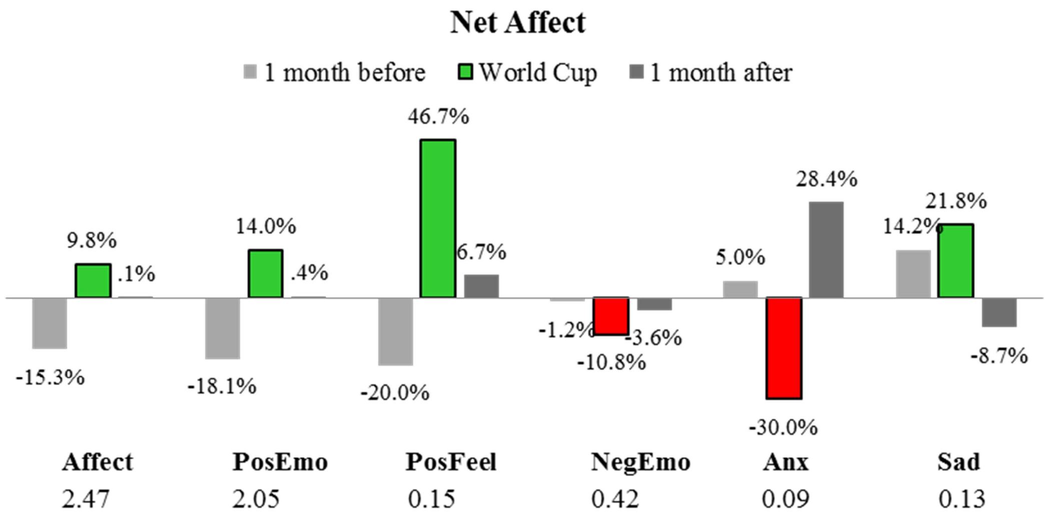


Figure 7.19 Net Affect changes during the World Cup to the aggregated (word count weighted) baseline of all scores before and after. All measures show relative changes, not absolute LIWC scores.

Additionally peaks occur for Social discourse (+13.9%) and first person plural pronouns (We) (+16.0%) (Figure 7.20). Seemingly, the World Cup increased aspects of communal belongingness along with making the community happier. Regularly singing national anthems, decorating houses and public viewing places with the general aspect of collectively being caught up in excitement about the sport performance seems to strengthen social relation ties in the KIT Facebook network. A confounding aspect exists with the category Inclusion (-2.3% during the World Cup). As noted in Section 3.2.3, the category Inclusion was mentioned as an indicator for belongingness. The relative dip could be due to the nature of sporting events and the discourse surrounding them: (e.g. “we won”; “they won’t defeat us”). This is unlikely to be the major driver though. While Inclusion is still negative relative to its baseline, it is less negative compared to the months immediately preceding or following the World Cup. A small uptick in Inclusion is seen during the World Cup, but it was too small to balance the other aspects of low inclusion in the KIT Facebook discourse.

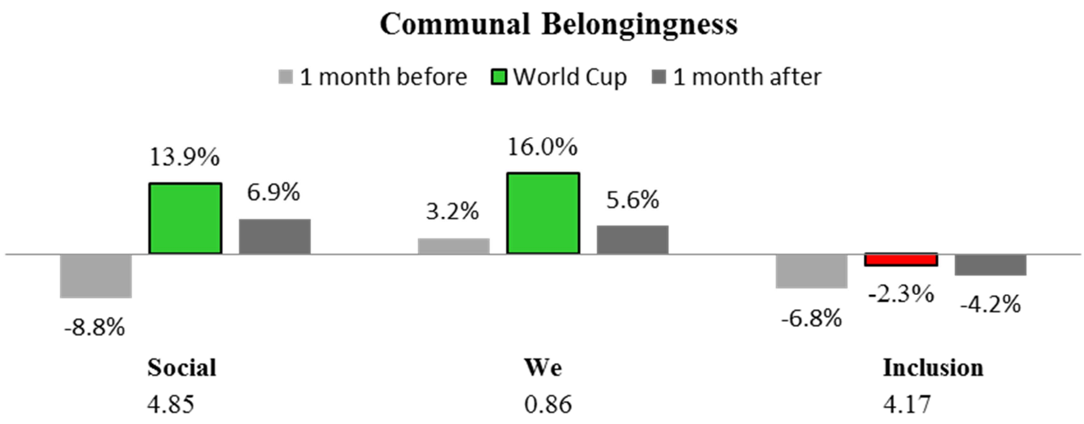


Figure 7.20 Communal Belongingness aspects during the World Cup to the aggregated (word count weighted) baseline of all scores before and after. All measures show relative changes, not absolute LIWC scores.

A more complex effect on the community was found in its time focus. Overall communication shifts event more to the present tense (+11.2%), suggesting a very day-to-day conversation across the network. Furthermore, there is an indication of a rise in self-confidence mirrored by the raise of Certainty expressions (+6.5%) relative to the rest of the semester. Finally, sentiment impacts of the World Cup are persistent overall. The sentiment increases and decreases in the consecutive month do not immediately return to the baseline but rather slowly decrease. This is a positive finding in light of the increases in well-being and communal belongingness.

7.3 Discussion

Focusing a Social Observatory on the KIT Facebook network revealed quite clear online discourse patterns among university network members. Post-comment comparison, in which posts represent activities of page administrators and comments participation of page visitors, serve as the sentiment analysis's baseline, providing both insights into the community characteristics as a whole, and as a guideline for further data partitions (**RQ 3**).

LIWC results display an overall satisfied community, disclosing indicators of high emotional and mental well-being through various emotional, attentional and cognitive categories. Interestingly, comments are both the most positive and negative aspects of the dataset, indicating that the community has a diversity of emotion even though the net effect is overall positive. In addition a general high level of communal belongingness is suggested by the high use in combination of inclusion words and social references, along with the low indications of strong social hierarchy. To better understand the dynamics of discourse, focus was shifted to differences between comments and posts, considering if it originated on an administrator or student led page. Comments on student pages are more emotional overall. Combining this with the prevalence of cognitive processes in comments, it can be posited that a central motivation for visiting the KIT Facebook pages is seeking lively discussions and discussion of opinions. In contrast, university administrators seem to restrict themselves to ‘newsflashes’ in a professional, formal manner, avoiding discussion.

Though the post-comment comparison is suited to gain first insights into the sentiment of the KIT community, it is with the partitioning of the database that communal attributes are uncovered. A dissimilarity analysis of Facebook pages evidenced that university topics have crucial impact on sentiment in communication. It was discovered the further a page was from other page in terms of sentiment usage, the less integrated into the KIT network it is (considering interaction distances on the social graph). High dissimilarity can be understood as tantamount to low relationship strength. Consequently, distance scores depict valuable information for pages to monitor their positioning within the community.

An overview of the literature benchmarks concerning discourse patterns estimated by LIWC is assessed in relationship with Section 3.2.3. Linguistic accommodation, the process of matching language styles of linguistic partners (Niederhoffer and Pennebaker 2002, 339), was assumed to be present due to its characteristics of post-comment exchanges found in Facebook. However its occurrence on KIT’s Facebook interactions is questionable. Generally high post-comment dissimilarity indicates that patterns of language style matching are not present. Exceptions are more likely attributable to individuals posting and commenting on different pages than linguistic accommodation. Although accommodation increases with more frequent interaction, a yearly analysis fails to support the hypothesis of language style matching of interlocutors.

An attempt to extract deceptive discourse from the KIT data was attempted. Four LIWC categories served as predictors of deceptive patterns as suggested by former research (Newman et al. 2003; Ott et al. 2011). Surprisingly posts contained a higher proportion of suspicious statements despite more severe consequences if untruthful statements are disclosed and an estimated higher detection risk. Student administrators show to be more inclined to use posts indicating deception than their university administrator counterparts. This tendency also holds true for comments. Subsequently, the more informal environment on student-run pages may reduce the visitors’ incentive to lie. Additionally, high deception scores for pages related to the page groups politics and sports were identified.

The way a Facebook page is administrated also seems to affect indicators contributing to well-being (**RQ 3**). Conversation on student-run pages tends to be lay higher focus on social interactions and is more concerned with individuals in the community. This indicates the existence of degrees of communal belongingness, especially on student-run pages. Whilst belongingness contributes to well-being (Huppert and So 2009) no administrative effect on emotive well-being is detected. This creates the condition of discussion staying on a more instructive or declarative level, which is not conducive to reflective, cognitively focused or emotive discourse. This leads to the secondary finding that communication is more homogeneous on pages administrated by students, with diminished emotive gaps between posts and comments as opposed to administrator-led pages.

With respect to the temporal aspects of the analysis, several interesting patterns were detected. Temporal dynamics illustrate powerful findings, contributing to the description of communal well-being. Campus discourse showed dependencies with the recurring semester cycles. KIT's Facebook community is most active when students are returning from holidays to the new semester. Accumulation of stress during exam weeks culminates in an overall negative sentiment valence through increasing anger, anxiety and negative emotion, as well as drops in positive affect. Supplementary pressure and study habits seem to reduce social activity in contrast to the middle of the semester, where social processes peak. The denial of the Elite status acted as a shockwave not only on the campus but also across the various pages of the university's Facebook community. Members emotionally reacted with anger, anxiety and sadness summarized by a generally increased density of negative emotion. Positive feelings in the community marked a significant drop in the week preceding the announcement. However, the community showed resilience as displayed by an increase in positive emotions three weeks after the event. Remarkably, the KIT community responded with an increase of communal belongingness to this disappointing experience. Finally, this analysis shows sensitivity to detection of internal and external events: The World Cup represents an external event with an emotional impact on the campus pages. Germany winning the World Cup displayed significant increases in net affect and communal belongingness, persistent even for a medium-term timeframe of a month past the awarding of the title.

7.3.1 Limitations and Future Work

Some limitations caused by the tools available do exist. As stated in the previous chapters, LIWC was not designed for short informal text like that found in Online Social Media, even though it copes astonishingly. A possible extension would be creating an additional dictionary with common abbreviations, phrases and emoticons that are pervasive in short, informal online texts as suggested by (Thelwall et al. 2010). Another necessary extension for the German dictionary is the splitting of formal and informal references to person. Otherwise it is not possible to accurately verify the level of formality in use across the community.

The importance of multilingualism in Online Social Media is increasingly recognized. Interlanguage comparison or even pages including a mixture of several languages could mislead interpretation of results. To allow for consideration of these inaccuracies further software versions could process an output reflecting word count percentages of contained languages. A more ambitious attempt in full automation may then even adapt each LIWC category based on the specific language deviations and the calculated proportion of content. However this requires in-depth analysis of crossover language patterns ideally based on Social Media content.

This work focuses on large spikes and dips with clear data signals in its current iteration. Innumerable smaller and unstudied incidents can add up and be responsible for large emotive shifts just as well as significant and sudden dips and spikes. This would be similar to predictively assessing the Davies J-curve (Davies 1962) based on short, informal data. Uncovering possible long-term predictors and data signals bears countless difficulties. This is due to the fast-changing features of and in social media, including strong dynamics without distinct attributes. The long-term analysis of events seems best suited for large-scale political interventions (e.g. (Böcking, Hall, and Schneider 2015)) or small and clear communities such as the KIT (Lindner et al. 2015).

A major limitation of this exploratory work is its reliance on estimations of emotional states. This is especially true for dictionary-based approaches that are insensitive to context or limitedly-sensitive and thus will frequently misinterpret ambiguous words and certain linguistic constructs as irony or sarcasm. Context-sensitive software is emergent and it is likely that newer versions of LIWC will include these improvements (Pennebaker et al., 2007). Although there is a high amount of agreement with established literature to indicate this study's validity, better grounding of the dictionary to context and not only latent states would allow for more definitive statements on the general health of the community. Envisioned in a full TSR system is a platform where both self-reported data and unstructured and informal texts like that on Facebook can be extracted and analyzed. In the long run it surely can be expected that this study's approach will benefit from fast developing improvements in sentiment analysis.

Some extension ideas for specific use case are possible. Former LIWC research has treated authorship characterization based on main characteristics as gender and age via selected tendencies for LIWC scores (Newman et al. 2008). The university use case could be suited to test the introduction of this feature to BeWell by testing whether sentiment tracked on pages for diverse study branches reflect the official KIT statistics on gender and age available for each study course. An interesting extension would be a comparative assessment of other universities and technical universities in Germany, as well as (dis)similar global universities. This would enable the establishment of in-depth comparisons of community characteristics and participative behavior, representing a powerful information resource for education institutions worldwide. It would also establish the findings this work as confirmatory rather than exploratory.

One major bias of utilizing Social Media text content to derive community characterization is the fact that there are a limited proportion of members who actively participate. Describing a given community with a Social Observatory therefore considers solely the members attracted to social media discourse. Thus community characteristics theoretically include the biases of restricted and incomplete member perception. Hence it is important to respect the distinction between online communities actually regarded and the complete community at which many findings aim. It is likely that relations and tendencies of the online presence are closely linked to the community as a whole, yet this conclusion cannot be made definitively. Meeting this problem can be best achieved by only carefully, if at all, generalizing results of active social media users to bigger parts of the community. This process should be made with consideration to each specific finding. For instance it is likely that the KIT online community's community-oriented reaction to the critical disappointment of the Elite denial is generalizable, whereas stating that people linked to university politics and sports show higher frequencies of lying would be an absurd generalization of Facebook specific discourse patterns.

In many areas this study was only able to execute first steps of completely envisioned capabilities and some possibilities have not been treated at all. Having delivered of the effectiveness of BeWell's attempt to community observation, it is hoped that further research will follow up this work. BeWell has provided first evidence that it is sensitive to sentiment peaks induced by short term events, external events, and time intervals. Calibration of these characteristics of events and time frames could allow for automated identification, further contributing to automation. Establishing highly sensitive signals to capture sentiment changes may reveal hidden influences on communities and is especially attractive linked with the possibility of real time data-feeds. Sometimes there exists severe interest in effects of events with focus on the incident itself, rather than aiming to describe the community by it. If the event depicts a macro level, affecting multiple communities, the Social Observatory can be adapted to extract short term databases of concerned communities and subsequently deliver a more complete picture. Policy impacts present just one of countless examples. Assessment of suitability for inter-community analysis in future research would extend the operational area substantially.

Discourse structure and preset rules differ enormously across social media and network platforms which has a distinct impact on tracked sentiment results. First work on these differences was approached in (Lin & Qiu, 2013). Empowering BeWell@KIT to track multiple social media platforms requires not only new functionality on the data extraction level, but will need similar information on sentiment baselines as provided by Lin & Qiu to allow for comparability over multiple platforms. Intensive combination and mutual enrichment of the two approaches, also referred to as Social Language Network Analysis, displays a whole set of additional approaches that could be addressed by future work (Scholand, Tausczik, and Pennebaker 2010).

7.4 Conclusion

BeWell@KIT has shown that it can detect notable community events by tracking expressed sentiment in Facebook posts and comments (**RQ 3**). Combining the stakeholder baselines with event-based tracking is interesting from a policy perspective, as it creates a communication mechanism for where stakeholders can present and discuss events and policy changes in a public forum. The contributions are twofold: this work binds a multi-dimensional well-being definition to publically available indicators that are otherwise hidden inside a data stream. To achieve this, both benchmarks from literature and unusual sentiment-based spikes and dips were observed and reported. Secondly this work is motivated by the university's desire to improve the understanding of itself as an institution. This work serves as a first attempt to develop and ground transformative services into the decision making process (**RQ 1.2**), with an aim to support member participation based on reliable information.

The results revealed by the temporal analysis indicate that within a community, stakeholders cannot be identified in a top-down way. Especially the shockwaves across the digital community after the loss Elite status show that the community is both self-nominated, and highly engaged, participating in the events and emotions experienced as a community. Partitioning the data in recurring semester cycles presents information on how communication focus shifts over the year. Due to the fact that people frequently debate about daily activities and events the results also capture the prevailing topics of daily activities. It was found that the stressful exam weeks lower emotional happiness while simultaneously show community members being less socially active.

Knowledge about such sentiment changes (cyclic and unexpected) may be put to use to advise feedback and community engagement attempts. For example, voluntary surveys might receive the highest participation at the beginning of the semester, when social processes peak and members show highest participation, instead of during demanding exam weeks. Similarly, detecting sentiment intervals such as semester cycles could advise when employees are most willing and able to put up with additional pressure, thus optimizing efficiency.

The way a Facebook page is administrated seems to affect a basic indicator contributing to well-being, namely the feeling of communal belongingness. This characteristic is especially valuable for institutions since it reflects if constituents can identify themselves with values and views of the organization. Sentiment scores showed ability to conclude the Facebook site's connectivity to other pagers when backed up with a Social Network Graph. Sentiment scores indicating (social) isolation could be passed to respective pages and evoke appropriate actions and research if this characteristic is pervasive. Whilst Social Network Analysis already provides this functionality, establishing integration levels through different data and sentiment analysis adds more depth. However, this possibility needs further evidence through matching future community results of sentiment scores and network graphs.

Part IV.
Finale

Chapter VIII Conclusion

"Achievement of your happiness is the only moral purpose of your life, and that happiness, not pain or mindless self-indulgence, is the proof of your moral integrity, since it is the proof and the result of your loyalty to the achievement of your values."

Ayn Rand (1905-1982)

Although it is well-known and accepted that everyone wants to improve their own well-being, a fully functional measurement system has yet to be introduced. The reasons are many, mainly due to outstanding complexities in the definition and identification of indicators of well-being, and their integration into social systems once identified. This thesis addresses these problems considering the forces of servicization, humanization, and digitalization of the modern economy. The increase in transparency caused by the rise of the internet increased individual's ability to compare and contrast their own lot, and demand services that support attaining the goal of being happier and more satisfied. Such services are called transformative services, or services that have the maintenance and improvement of individual and communal well-being as a goal function (Anderson et al. 2013; Rosenbaum et al. 2011). The movement to transformative services inclusive of human well-being necessitates the formalization of a method to define and identify well-being, measure well-being, and evaluate the characteristics thereof.

Following Service Dominant logic (Vargo 2009), this thesis evaluates two applied methods for the measurement of well-being considering digital fora: gamification and text analytics propagated on the social media and networking platform Facebook. As the definition and determinants of well-being and happiness are of the upmost importance for a successful human-oriented service, the first emphasis of this thesis was in establishing how well-being is defined and experienced. In the second section this thesis concentrated on the unobtrusive detection and evaluation of well-being gained from short, informal text harvested from Facebook posts and comments. In particular, this thesis focused on bias-free methods of social media analysis, tested on multiple independent use cases.

Section 8.1 summarizes the contributions of this thesis by addressing and appraising the Research Questions of Chapter 1. Section 8.2 critically discusses the assumptions and limitations of this work, and closes with an overview of future work.

8.1 Contributions

This thesis focused on the definition, refinement, and application of well-being as a progressive community management service for use in institutional settings. Its contributions to the TSR literature and service research community more broadly are threefold:

> 1) The design of a multi-tiered service framework as a means to estimate the entirety of the service environment as it pertains to well-being,
>
> 2) The technical implementation of a data extractor as complementary methodology to study such systems,
>
> 3) The understanding of relevant indicators of the evaluation of personal and institutional well-being.

Particular care was taken to consider design requirements and their impact on the application thereof. The three contribution aspects are discussed in more detail in the subsequent research questions.

8.1.1 Defining Well-being for Transformative Service Research

There is near universal agreement that everyone deserves to be happier and that individuals' well-being is paramount for healthier, happier communities. What has not yet been agreed upon is how to define (in the first instance) and then measure (in the second instance) that which is essential to well-being. These two aspects are critical. Without a reliable definition and measurement, metrics based on well-being or happiness cannot be elevated past the normative. However with a clearly defined and consistent metric system, well-being is poised to become an invaluable metric in the effort to humanize the modern economy and service ecosystem. Due to these interdependencies this thesis focuses first on a comparative analysis of the major well-being definitions and measurements. This was the motivation behind Research Question 1:

RESEARCH QUESTION 1.1 ⊰ DEFINING WELL-BEING ⊱ *Which attributes of well-being's conceptual definitions allow for the operational usage of well-being in institutional management?*

The first step to address this research question is to analyze the requirements for capturing normative states in order to determine different service layers. As the definition of well-being lacks a *fil-rouge*, Research Question 1.1 distinguishes the necessary attributes and identify relevant aspects of a singular well-being definition. It is necessary to measure the positive and negative feelings of the experience of well-being; it is also necessary to recognize that the aspects which afford pro-social and thus pro-institutional well-being are not always consistent with being happy all of the time. Here the conversation changes from ex-post measurements to the difference between being happy and satisfied. This thesis finds that both metrics are necessary for a complete institutional measurement.

As such, this research advocates the operationalization of the tri-layered approach Human Flourishing (Huppert and So 2013), with its concentration on positive emotions and positive characteristics (individually conducive to well-being) and positive functioning (communally conducive to well-being). Chapter 2 further contributes a formal notation of Human Flourishing (Equation 2.3) as by prioritizing the experience of positive emotions while implementing that all constructs are necessary to being well. In the case that the construct positive emotion or two items from either positive characteristics or positive functioning are not present, the individual is considered to be not flourishing.

It must be recognized that Human Flourishing is still merely a marker of temporal well-being, meaning that it is the weather and not the climate that is identified. In order to more realistically ‘estimate the climate’ it is necessary to review even more personal psychometrics, namely personality. This thesis established that two personality types, extroversion and neuroticism, are responsible for between 54-70% of an individual’s perception of well-being. Thereby, Research Question 1.1 investigated not only the temporal estimates of well-being, but also the foundational determinants of well-being. Accordingly in the process of addressing Research Question 1.1, Chapter 4 establishes on the relationship between well-being and personality.

Applied methods - even if developed for big data assessment - reveal interesting and new facets of this study's well-being prediction problem upon comparably small datasets (Chapter 4 and Appendix II). Social data availability simplifies the understanding of dependencies and underlying structures, but it will also demand for easy-to-use, well-interpretable, but nevertheless powerful analysis procedures. The topic of 'small data' analysis including small samples with high dimensionality recently evolved from increased availability of individual, personal data gained for example from smart phones and social media activity. It is consequently proposed that non-parametric tools and feature selection methods should be further developed and more often be utilized in order to question popular, but simple regression results. Applied non-parametric machine learning algorithms significantly increased the developed picture of the well-being dependencies' internal structures. Today, most analyses on social problems do not challenge significances found by variance analysis and linear regression for underlying non-parametric structures, although those would probably add additional value to the ongoing scientific discussion.

RESEARCH QUESTION 1.2 ⊰ TRANSFORMATIVE SERVICE RESEARCH ⊱ *What are the necessary attributes for constructing well-being oriented service design for institutional management?*

Transformative Service Research is essentially service research where the well-being of the entire service value chain is maintained or increased. To implement such a worthy design, the interaction of well-being and services must be mapped. This thesis contributes to service research a framework which is at once reflective of the individual and networked across the various services that impact individuals on a day to day basis. This is achieved with the introduction of a tri-layered framework that considers macro-, meso-, and micro-level interactions between individuals and services.

Respecting the value of tangible and economic assessments of well-being, an assessment paradigm for the design of services must retain a macro assessment of the environment or ecosystem in which the service is expected to be deployed. Trivially explained, the 'day to day' of an average citizen of the Democratic Republic of Congo and Norway are different and must therefore are for in transformative services. Moreover, the critical relationship of an individual and their immediate environment must be considered. This meso-perspective is similar in breath to the concept of Human Flourishing: individuals work in an environment, and that interaction is a key part of their perception of well-being. One particular meso-characteristic, 'belongingness,' is established as a key indicator when assessing institutional well-being unobtrusively. With belongingness one simultaneously estimates the meso-perception, and gains insights into micro, psychological aspects of well-being. This estimation of belongingness is the foundation of Research Question 3.

As established by Research Question 1.1, this is only part of the picture when it comes to well-being. Transformative services must likewise consider the individual's psychological profile. The interplay between extroversion, neuroticism, and an operationalized Human Flourishing corresponds well to literature-based benchmarks in happiness research (Chapter 4). This micro-aspect has been heretofore untouched due to myriad ethical, legal, and practical considerations including scalability. In the era of Big Data, the ability to analyze exactly this micro-consideration has changed. As such this work contributes to that research gap.

Missing is an application that can extract this information in a privacy friendly and scalable way. This is crucial as before each aspect has been considered, a realistic and functional transformative service cannot be designed, for it is within this networked, layered environment that the cycle of service provision, perception, and influence take place. Even more importantly, it sets the stage for information-driven transformative service design. Research Questions 2.1-2.4 address these necessary aspects of empirically-based transformative service design.

8.1.2 Refining the Data Characteristics of Digital Well-being

In addition to difficulties in pinpointing the measureable attributes of well-being, virtually unknown is the suitability of such data. Previous efforts in well-being assessment tend to be longitudinal studies based on survey responses, with measurements taken at infrequent intervals. Necessary for an institutional level assessment is shorter, more frequent intervals nearing real-time reporting of constituent well-being. This leaves the open research and design challenge of formulating well-being assessment in such a way that it can be either pushed to constituents frequently or pulled from constituents at predefined intervals and granularities in a way that is robust and reliable.

RESEARCH QUESTION 2.1 ⊰ DATA HARVESTING ⊱ *Considering the methods gamification and text analytics in online social media, which method is more appropriately applied to extract near to real time well-being data in a continuous manner?*

Considering Research Question 2.1, through these exemplary case studies, it is obvious that the use of text analytics and the related sentiment analysis to evaluate human well-being in terms of Human Flourishing provides a more holistic and robust method of analysis. The first case study exemplified that gamification is a meritorious approach though it suffers from several context dependencies. With such an approach, data extracted is truthful and personal. However, the method struggles with issues of participant fatigue. The second use case demonstrates superior facilities in the extraction of well-being data without participant fatigue or researcher bias. In addition, text-based approaches can be easily split along a variety of granularities, allowing for different community perspectives to be taken into account.

In concentrating on the platform Facebook, significant efforts must be deployed in verifying the findings from existing literature. Results cannot be considered reliable or valid when changing the data elicitation medium without an additional verification step. As Facebook is a relatively closed platform for quantitative studies, this is currently a research lacuna. Research Question 2.1 contributes to exactly this problem: verifying existing relationships from literature with two applied methods sourcing Facebook data.

Gamification has the merit of reproducing known relationships with a validated method. It has other drawbacks (addressed below) that make it a prohibitive mechanism for large-scale studies. Text analytics, whist certainly not without its own limitations, was found to be the more promising mechanism for the estimation of well-being in digital communities. Text committed to the public pages of an online social media platform like Facebook is granular, constantly updating, highly individualized, and carries latent aspects of personality. In the case of public pages (such as in Chapters 5 and 7), it is also freely available but does not carry aspects of research design bias. In the case of requested data from individual pages (i.e., Chapter 6), researcher bias is by and large mitigated by the use of Facebook, where data is granted in its entirety. The setup of the extraction process per Facebook's regulations means that participant fatigue is out of scope in such a design. The analysis of such short, informal text is well-done with a dictionary-based approach like that found in the Linguistic Inquiry and Word Count package. This limitedly context sensitive toolkit extracts word frequencies given a sentiment category, giving researchers a mechanism to estimate language and emotive patterns with a common baseline. As the tool concentrates on how language is used, rather than what is being said, it also supports this thesis's aim in measuring the climate and not the weather.

Having identified three layers of service requirements and the need to extract potentially sensitive data in these stages, as a next step this knowledge can be applied to design technical solutions by way of an information-driven TSR application. To exemplify the usefulness of the information-driven approach, Part III presented case studies on two methodologies. The aim of the first methodology was to study the effects of gamification on incentivization and participation, as addressed by RQ 2.2; the second methodology investigated the suitability and attributes of text analytics for unobtrusive detection of well-being (RQ 2.3).

RESEARCH QUESTION 2.2 ≺ GAMIFIED SURVEYS ≻ *Does the gamification of surveys enable frequent, granular views of individual's well-being without a high participant drop-out rate?*

Gamification can be successfully applied to measure Human Flourishing – motivating users to continuously employ the artifact while providing truthful data. However, serious gaming for well-being revelation has some serious limitations and conflicting results for some incentives. Chapter 4 establishes that the primary interest of the users was to calculate and track their HFS, and to investigate their Flourishing constructs. Participants predominantly liked the gameful approach. Social incentives and exchanges built into the platform were underutilized, supporting the view that the users prefer their well-being information to remain self-contained. There is an observable rejection of comparative and evaluative incentives through users with high(er) neuroticism.

While participants were satisfied overall with the approach and gamified approach, two major limitations were self-evident. One is the high level of self-selection bias and reference effects incurred. As a realistic estimate for *n* possible participants on the online social network Facebook is not possible given the limitations of the platform, it was not possible to create a bias-mitigating variable from which to test the reliability of the results. This is a serious consideration for researchers intending to gamify personality and well-being surveys. The second necessary consideration is participant fatigue. An observable drop in participation occurs after approximately four interactions with the game. Gamified personas are not identified as a deterrent to the collection of well-being data in serious games. This indicates that while granular and truthful information can be extracted in a serious game, the frequency of data-extraction does not fulfil the requirements of a transformative service.

RESEARCH QUESTION 2.3 ⊰ RELIABILITY AND VALIDITY ⊱ *Which well-known relationships between well-being and personality can be reproduced when using text-based data found in social media posts?*

Social media data has several very specific characteristics that make verification difficult. Resources like Facebook posts and comments are relatively short compared to more traditional, non-digital corpi. Validation on small data is a well-known methodological issue. Due to the brief and informal nature of such resources, abbreviations and slang that are commonly used and broadly know yet rarely committed to a dictionary are frequently used. Considering the degree of such usage, quite a bit of latent emotive data could be lost due to recognition issues.

These methodological challenges notwithstanding, the combination of Facebook data and LIWC analysis applied in this work has proven to deliver reliable, valid and robust results. In what could be a particularity of the German user sample, overall use of fillers, slang and other non-fluencies averaged at under 0.01% in all samples. And while individual posts are short, the aggregation methods applied in Chapters 5, 6, and 7 allowed the overall corpus per instance to be large enough to allow for validation. As a robustness check, this thesis varied the minimum amount of words per analysis and found that the results do not change significantly when at least 50 words are recognized by LIWC's internal dictionary. This is a significantly lower threshold than in previous works.

LIWC analyses across Chapter 5, 6, and 7 revealed strong relationships with the constructs found in Human Flourishing (positive emotions, characteristics, and functioning). Moreover, aspects of communal belongingness were identified and analyzed in Chapters 5 and 7, helping to identify the overall well-being of the institution and not only the individual. However, the initial analysis found in Chapter 5 also suggests that there are some issues of establishing ground truth. Initial attempts to find a relationship between the personality factors extraversion and neuroticism and LIWC's positive and negative affect categories could not be verified. This confounding result is the basis of Research Question 2.4: to which extend is the medium affecting data quality, and can these effects be identified and later mitigated?

RESEARCH QUESTION 2.4 ⊰ DATA VERACITY ⊱ *Are discernable characteristics of active representation identifiable, and if so, what are these characteristics?*

Just as well-known as it is that people are multi-faceted, it is well-known that individuals pick and choose aspects of their activities and personality to alternatively highlight and censor in a given forum. A trivialized example is that when speaking to one's boss and about one's boss with a spouse, the tone and content of such a conversation can and will change. Unknown is how this instinct displays itself across online social media, and if there is an impact on the data of this in the first place. Literature is inconclusive, and has been poorly assessed on a per-platform basis. This thesis, and specifically Chapter 6, addressed this Facebook-oriented research gap by aligning Facebook posts and self-reported survey data on self-representation and Facebook usage, along with personality data.

The findings of Chapter 6 confirm that self-presentation is indeed a phenomenon that exists, and it has an impact on the way that LIWC's internal algorithms process data. Chapter 6 also finds that aspects defining the degree to which one self-represents is identifiable. As it is identifiable, it is mitigatable. To reliably mitigate the impact of self-representation, this thesis first establishes the categories of LIWC with highly significant relationships to personality factors. It then clusters those factors to assess the 'personality' of a text corpus. Applying this method not only mitigates self-representation in Facebook analyses, it also identifies the baseline of individuals' personalities. This is by extension a contribution towards quantitatively establishing ground truth from Facebook data.

8.1.3 Applying Transformative Services

Service design is transformative when it has a measurable, even optimizing, positive affect on human well-being. Any prospect for such felicitous outcomes, however, requires accurate assessment or measurement of well-being in and for target populations. Such assessment raises two immediate issues: conceptualization (How should well-being be conceptually operationalized?) and measurement (Given an operationalization of well-being, how can it be measured?). This was addressed by Research Question 1.1. Implicit in the tri-layered definition of well-being and its dependency on psychological aspects of personality are the first aspects of transformative service requirements. Research Question 1.2 uncovers and delineates these attributes as they pertain to transformative service research.

RESEARCH QUESTION 3 ⊰ CHARACTERISTIC MAPPING ⊱ *Can community characteristics like well-being and organizational belongingness be unobtrusively established? If so, what are the key characteristics?*

Sentiment analysis of German Politicians on Facebook and the KIT Facebook presences revealed multiple characteristics useful to describe a community, facilitated by the technical solution named a Social Observatory. LIWC score interpretation allowed for the community's well-being, communal belongingness, emotionality, formality level and honesty to be established. The description of characteristics was not restricted to capturing macro tendencies but even delivered dynamics over time, sentiment cycles, and differences between various subgroups of the respective community. Results affirm LIWC as an efficient analysis tool for tracking communal sentiment, well-being and aspects of belongingness. It is found that LIWC categories related to emotional affection, attentional focus (i.e. pronoun use) and cognitive and social processes were especially crucial to derive the central findings. The results are quite often nuanced: small percentage points highlight differences for more than one community characteristic. Yet, topic domains and specific other scores allow for detecting more specific interpretations and should not be disregarded. The list of all LIWC categories deployed as well as the volume of words used in a given setting gives a wide and holistic impression of guiding characteristics. One interesting caveat to this Research Question 3 is its dependency on word count. This thesis tested then employed a cut-off of 50 words for psychometric analyses and 34 words (the average German sentence length) for line by line analyses. While this is well below the thresholds of similar studies (Yarkoni 2010; Berber-Sardinha 2000; Sheridan-Dodds and Danforth 2010). Pages, posts and comments below the employed thresholds cannot be considered, and if they are subject to serious considerations of validity and reliability.

Information estimated from aggregated social media data may lack some interpretation quality but provides an easy and repeatable way to gain quick insight into the essential factors defining a community. Macro-assessment of social indicators rises from investigation of post-comment distinction, a pre-given structure of any Facebook dataset. This means that the approach is easily replicable for other communities and generalizable. Although some customizing effort concerning data preparation are inevitable if community-specific insights are pursued, many of the employed partitions are to be individualized to further use cases. This aspect of popularly sourced well-being information is ripe for adaptation into transformative service research. By utilizing this multi-faceted picture of the individual, BeWell@KIT as implemented with the Social Observatory encourages communities to proactively manage the components causing well-being (or its counterpart, ill-being) as a form of adaptive people management. Through the observation of a decrease in well-being, participatory approaches to decision making and policy making could be applied as a means to reengage previously content constituent-users, and engage new constituent-users throughout the community.

The workflow of the Social Observatory (exemplified in Chapter 5 and Chapter 7 as BeWell@KIT) equips social researchers with a new way to unobtrusively select, analyze and compare communities of interest in a highly automatable surrounding. As institutions seek to evoke participative interaction with stakeholders, learning about the driving forces of participative behavior is the foundation to further induce frequent feedback of members on the social media platforms but could even be beneficial to participation via other media. High participation can not only function as an effective measure to reveal the reasons behind eventually to be detected well-being drops in the future, but has shown to positively influence happiness of communities (Frey and Stutzer 2001).

8.2 Research Outlook

This section addresses the limitations of the thesis and suggests areas of future work. The integration of well-being into service design is in its infancy. While important questions on the operationalization thereof have been addressed in this thesis, several aspects remain underdressed. These areas have promising research value, and may provide valuable insights in the future.

8.2.1 Technical Considerations in Transformative Service Research

Transformative Service Research is poised to greatly benefit both the academic and the practical aspects of the service economy. This thesis points out areas of further technical developments that should be pursued in order to fully integrate TSR into the digital economy. These are discussed below.

Further Integration of Mechanisms

This thesis investigated two applied methodologies for the extraction of well-being data from digital platforms: gamification and text analytics. Where the application of serious games to survey data had the positive attribute of individually sourced and thus the most accurate data, it also had high participant fatigue. Text analytics is an estimation of ground truth, but can be extracted with any time frame as it is user independent. Yet to be addressed is the combination thereof. An interesting method to be investigated would be the extension of a platform as introduced in Chapter 4 to include streaming text analytics capabilities. This would decrease the necessary amount of pulled questions from participants while still maintaining the granularity of text analytics. Integration is chiefly a design issue, and would benefit from the application of design science (Hevner 2007). Design science would also facilitate the creation of a well-being dashboard from which progressive community management can be directed. Interesting future directions also include the impact of interacting with such a system on socially responding or social desirability aspects from the perspective of Common Method Bias (Podsakoff et al. 2003).

Learning Approaches in Digital Discourse

This thesis chiefly concentrated on the measurement of psychometrics as a predictive aspect of well-being. Machine learning approaches were applied to gamified survey data in an effort to predict well-being of individuals In Chapter 4. A similar tactic could be applied to the text-based data in order to discover not only the latent values of the words used, but the topics within them. Where this work concentrates on applying dictionary-based counting algorithms, machine learning methods suitable for unstructured data, generally called topic modelling, including n-grams (Oberlander and Nowson 2006) and Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003) can also be applied.

In particular the work (Youyou, Kosinski, and Stillwell 2015) suggests that an unsupervised learning approach can predict personality. They were however unsuccessful in predicting happiness. This leaves the open question of using personality to predict well-being using an exploratory language modeling approach. This unobtrusive approach is an ethically superior method to that of (Kramer, Guillory, and Hancock 2014), as shown in the work (Coviello, Fowler, and Franceschetti 2014). Another interesting aspect is using linguistic patterns to identify writing consistency in order to better identify temporal aspects of well-being (Runge et al. 2012; Argamon et al. 2009).

Cross-Platform Validation

Consistent with literature on self-presentation (Hogan 2010; Special and Li-Barber 2012; Lingel, Naaman, and boyd 2014; Mehra, Kilduff, and Brass 2001; Lin and Qiu 2013) as well at the principles of validation in research, it remains to be addressed what parameter changes (if any) must be applied when estimating well-being from other online social media platforms. First work in the comparison of the same use on different platforms has been addressed in (Lin and Qiu 2013). Of particular interest to community managers for further validation could be enterprise social networks from the perspective of the well-being of professional institutions, and the closely linked online social networks which specialize in professional networks like LinkedIn or the Germany-based Xing platform. Whereas an enterprise social network would have special considerations due to privacy concerns, the interesting aspects of platforms like LinkedIn is the scarcity of words used as well as words allowed in a profile.[45] Restricted word counts present an interesting validation challenge considering the overall small n in use (Braga-Neto and Dougherty 2004). A similar small-n challenge exists for low volume users of the micro-blogging platform Twitter. Visually-based social networks like YouTube, Snapchat, and Pintrest are also of interest considering their growing user bases. Especially wtih self-representation, there are considerable research gaps. However, the technology behind machine vision that would be required to classify such aspects is unfortunately still lacking (Poczos et al. 2012).

8.2.2 Human Factors in Institutional Management

In addition to technical considerations in Transformative Service Research, two aspects of human interest should be further addressed: digital research ethics and the use of such data in participatory decision-making.

[45] https://www.linkedin.com/pulse/20140319195712-109230363-linkedin-maximum-character-counts-for-2014

Ethical Considerations in Digital Communities Research

An underdressed aspect of digital communities' research is informed consent. The Terms and Conditions across social networking and social media platforms are unanimous: that which is committed to the platform can and will be used in research. Simply put: registration on or with the platform indicates continued agreement with this statement. Even more concerning is that resources which are committed publicly are considered a part of the public domain if the user understands this or not. Comprehension is taken for granted, though it has been shown that the Terms and Conditions are often written in legal jargon far above the average reading level of participants (Fiesler and Bruckman 2014). While this thesis conformed to the Terms and Conditions of all utilized platforms in addition to following the guidelines of (Markham and Buchanan 2012), it remains an open question if informed consent can and should be maintained inside of user Terms and Conditions from the perspective of user assent and user comprehension.

Several aspects come into question, with the foundational question being if community members consciously understand that agreeing to Terms and Conditions is implicitly agreeing to Informed Consent as well. As seen in the controversy surrounding the Facebook study on emotional contagion (Kramer, Guillory, and Hancock 2014), this assumption is questionable at best and should be addressed by the research community. Working from the assumption that participants do in fact know that their data is considered a valid research source, the next research issue is if users understand the extent of data which they agree to grant researchers on social media platforms. This has wide-reaching implications, from personal information, the information of friends, to intellectual property rights. A knowledge-based experiment of the permissions and boundaries of users on social networks should be conducted for this purpose. Additionally, a stronger ethically-based research guideline should be issued in cooperation not only with academics but also with the platforms themselves for digital research and researchers, consistent with the proposal of (Friedman, Kahn Jr., and Borning 2003; S. H. Schwartz 1994; Friedman 1996).

Participatory Decision Making

The overarching goal of deliberative participation procedures is yielding user-generated debates and results on complex topics. Participation behavior has changed a lot in the era of digitalization (Boulianne 2009). That which were previously considered obstacles, such as time and space, are decreased and simplified by digital participation in political, as well as in corporate or private contexts. This is especially true for young(er) institutional constituents, though not exclusively (Escher 2013; Hampton et al. 2011). This development has affords the ability to change public management dialogue from a uni-directional flow from the institution to users into consultative or participative bi-directional flows between users and the institution (OECD 2007; *OECD* 2010). This is a positive development but requires further academic studies on participant motivation and incentivization (Haas, Caton, and Weinhardt 2011; Margetts et al. 2011; Bishop 2007); and user-oriented design principles (Friedman, Kahn Jr., and Borning 2003; Larsson et al. 2005); theoretical and applied participation tactics (Dworman, Kimbrough, and Laing 1995; Zhong, Kimbrough, and Wu 2002; Vassileva 2012).

This thesis provides first evidence that a digital tool which is sensitive to sentiment peaks induced by short term events and time intervals can be applied in progressive community management. This advances the literature surrounding Transformative Service Research. The next step is creating an automated sentiment feedback tool for use in participatory decision making. A deeper understanding of the emotional motivation behind online participation behavior is inevitable to improve the user friendliness and experiential aspects of participatory platforms. Personalization simplifies the use of such platforms and keeps the user motivated to participate. Envisioned is an open dashboard fed by Facebook and other feeds. This can be used to highlight community mood and might, combined with advanced learning techniques, lead the users through the platform depending on their personal current mood. Therefore the participatory interaction within the group is facilitated. In support of institutional efforts this anticipates a happier, healthier community.

Part V.
Appendix

Appendix I Survey Items Considered in Part III

Five Factor Inventory:

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who *likes to spend time with others*? Please write a number next to each statement to indicate the extent to which **you agree or disagree with that statement.**

I am someone who…

BF1. _____ Is talkative

BF2. _____ Tends to find fault with others

BF3. _____ Does a thorough job

BF4. _____ Is depressed, blue

BF5. _____ Is original, comes up with new ideas

BF6. _____ Is reserved

BF7. _____ Is helpful and unselfish with others

BF8. _____ Can be somewhat careless

BF9. _____ Is relaxed, handles stress well.

BF10. _____ Is curious about many different things

BF11. _____ Is full of energy

BF12. _____ Starts quarrels with others

BF13. _____ Is a reliable worker

BF14. _____ Can be tense

BF15. _____ Is ingenious, a deep thinker

BF16. _____ Generates a lot of enthusiasm

BF17. _____ Has a forgiving nature

BF18. _____ Tends to be disorganized

BF19. _____ Worries a lot

BF20. _____ Has an active imagination

BF21. _____ Tends to be quiet

BF22. _____ Is generally trusting

BF23. _____ Tends to be lazy

BF24. _____ Is emotionally stable, not easily upset

BF25. _____ Is inventive

BF26. _____ Has an assertive personality

BF27. _____ Can be cold and aloof

BF28. _____ Perseveres until the task is finished

BF29. _____ Can be moody

BF30. _____ Values artistic, aesthetic experiences

BF31. _____ Is sometimes shy, inhibited

BF32. _____ Is considerate and kind to almost everyone

BF33. _____ Does things efficiently

BF34. _____ Remains calm in tense situations

BF35. _____ Prefers work that is routine

BF36. _____ Is outgoing, sociable

BF37. _____ Is sometimes rude to others

BF38. _____ Makes plans and follows through with them

BF39. _____ Gets nervous easily

BF40. _____ Likes to reflect, play with ideas

BF41. _____ Has few artistic interests

BF42. _____ Likes to cooperate with others

BF43. _____ Is easily distracted

BF44. _____ Is sophisticated in art, music, or literature

Human Flourishing Scale:

HF 1.*Competence*

Most days I feel a sense of accomplishment from what I do

HF 2. *Emotional stability*

(In the past week) I felt calm and peaceful

HF. 3 *Engagement*

I love learning new things

HF 4. *Meaning*

I generally feel that what I do in my life is valuable and worthwhile

HF 5.*Optimism*

I am always optimistic about my future

HF 6. *Positive emotion*

Taking all things together, how happy would you say you are?

HF 7. *Positive relationships*

There are people in my life who really care about me

HF 8. *Resilience*

When things go wrong in my life it generally takes me a long time to get back to normal.

HF 9. *Self-esteem*

In general, I feel very positive about myself

HF 10. *Vitality*

(In the past week) I had a lot of energy

Facebook Usage:

SM1. How often do you log into Facebook?

SM2. How often do you update your profile?

SM3. How many Facebook friends do you have?

SM4. Who are you interested in contacting on Facebook?

SM5. What do you find yourself frequently "Liking"?

SM6. Do you leave your contact information (Email, phone number, address) public on Facebook?

SM7. Which information about yourself do you have available on Facebook?

SM8. To which degree do you agree with this statement? “People should present themselves on online social networks as the same person as they are offline.”

SM9. With which of the following statements do you agree? (Choose all that apply)
I use Facebook …

A. because contacting to others is simple

B. because I'm curious, about the kind of life of people I do not know

C. to be recognized by others

D. because I can observe people around me

E. to obtain support from others

F. because I can learn a lot about others without me having to be seen

G. to inform others what I'm doing

H. to show everyone what I know and what I can

I. because this is how people connect nowadays

J. because I can reach many people

K. to give something and, if necessary to get something back

L. to show a different side of myself

SM10. Do other people present themselves differently in online and offline settings?

SM11. Complete the following statement. I manage my image on Facebook with (Choose all that apply)

A. group memberships

B. personal interests

C. a profile picture that shows my face

D. likes

E. my Friend List

F. a profile picture that is not obviously me

G. Albums

H. my Cover photo

SM12. Do you upload pictures to Facebook?

SM13. Other people represent themselves on Facebook by ….

A. group memberships

B. personal interests

C. a profile picture that shows their face

D. likes

E. Friend List

F. a profile picture that is not obviously them

G. Albums

H. Cover photo

SM14. To what extent do you agree with the following statements?

A. I quickly understand how I am perceived by others.

B. I can determine myself what I do or do not show others.

C. I can show personality completely.

D. I can be who or what I want on my Profile Page.

E. I can be more open online than in real life.

F. Online, I can present myself to everyone.

Appendix II A Comparative Assessment of Machine Learning Algorithms for Well-being Assessment

2.1 Kernel-Smoothing algorithms

The following kernel-smoothing algorithms are applied to solve the general prediction problem including the per-participant averaged HFS as dependent variable and the 13 demographic and personality variables as predictors. All variables are normalized to zero mean and SD one.

2.1.1 K-nearest neighbor

The introductory kernel method is a uniform kernel, including the k-nearest neighbors of the requested point into the analysis. For the k-nearest neighbor algorithm the dependent variables' value of these k neighbors within the training set are averaged. In R the algorithm is implemented using a knn package.

The implemented algorithm allows for an adjustment of the metric, by which the distance for k-nearest neighbors are calculated. By using the Minkowski distance the 1_{1-} (Manhattan-) and 1_{2-} (Euclidian-) metric and graduations in-between can be applied through a distance parameter (1 for Manhattan and 2 for Euclidian metric). Furthermore, differing kernels including Gaussian, Epanechnikov and the standard uniform, also referred to as rectangular kernel, are applied and compared.

The results show a slight superiority of the Euclidian metric for all kernels, why the l_1-metric is not further considered (Figure 8). The prediction accuracy is best for the Epanechnikov kernel *at* $k = 22$ $(RMSE_{Epan.} = 0.792)$. The Gaussian and uniform kernels perform best for $k = 12$ $((RMSE_{Gaus.} = 0.794$ and *(RMSE_{StrUni.} = 0.796).* Figure 1 provides a graphical representation. Nevertheless, all results are significantly worse than the *GLM (RMSE = 0.678)*. The given results already indicate that a static local structure might not be present within the data.

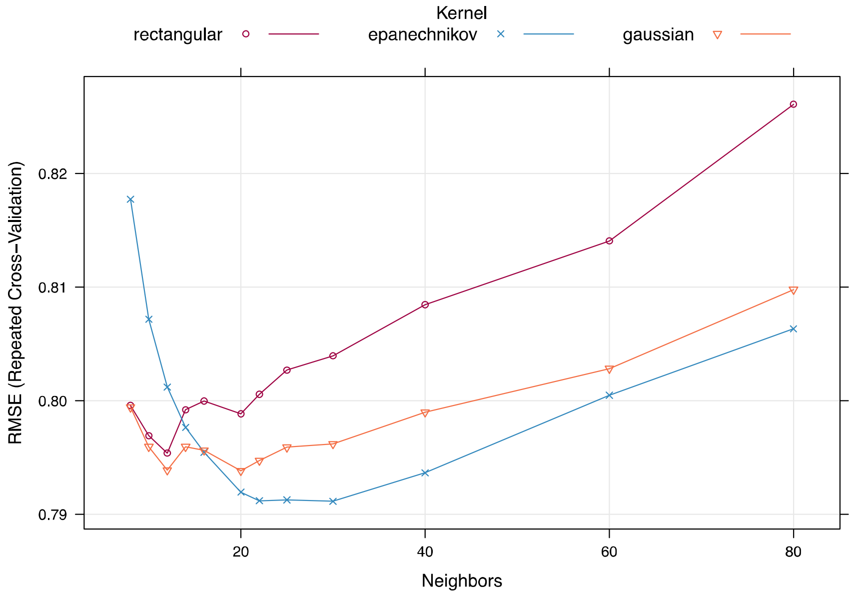


Figure 1. RMSE for k-nearest neighbor using Euclidian metric

However, the importance of the variables differs from the GLM's variance importance. As seen in Figure 2 neuroticism gains even more importance, while the demographics lose influence on the independent variable HFS.

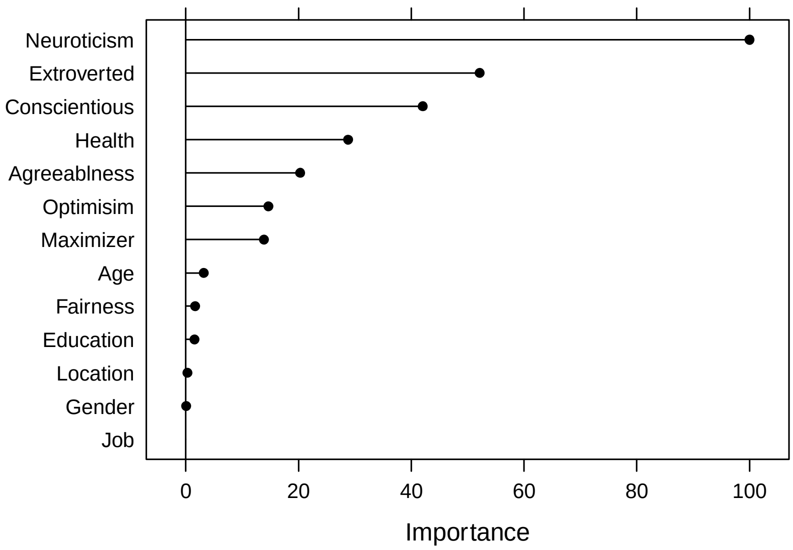


Figure 2. Variance importance for k-nearest neighbor using Euclidian metric

2.1.2 Non-parametric Regression

Non-parametric regression refers to algorithms, which calculate a local linear regression within a kernel environment instead of averaging the nearest neighbors. Three different non-parametric regression algorithms have been tested, namely an *Generalized Additive Model* *using LOESS*, a *Generalized Additive Model using Splines* and *Nonparametric Regression* (see Hayfield and Racine 2013).

2.1.3 LOESS

The LOESS (locally weighted scatterplot smoothing) algorithm fits a linear or quadratic regression within k-nearest neighbor environment with a uniform shape. The kernel's size is defined by parameter α, the proportion of training data points included in each kernel. For α = 1 all training points are included in every kernel, while α = 0.25 takes the 25% nearest points of the entire training data into the kernel. LOESS consequently turns into a GLM for α = 1. The distance calculation for the neighborhood definition is conducted with the tri-cube weight function: $(1 - (distance / max(distance))^3)^3$.

The algorithm is implemented using the caret package's gamLoess model. GamLoess implements the LOESS algorithm separately for each independent variable within a *Generalized Additive Model (GAM)*. Due to high computational costs, only the linear regression has been conducted. As seen in Figure 3 the accuracy converges towards the GLM's accuracy at 0.678, when α is close to one. However, an increase in accuracy cannot be observed when α is reduced. This result is in line with the previously mentioned low accuracy of the k-nearest neighbor algorithm. Noticeable is the RMSE drop for α = 0.32, which equals approximately 103 training points included in the local regression. This configuration does not outperform the GLM (*RMSE = 0.753*).

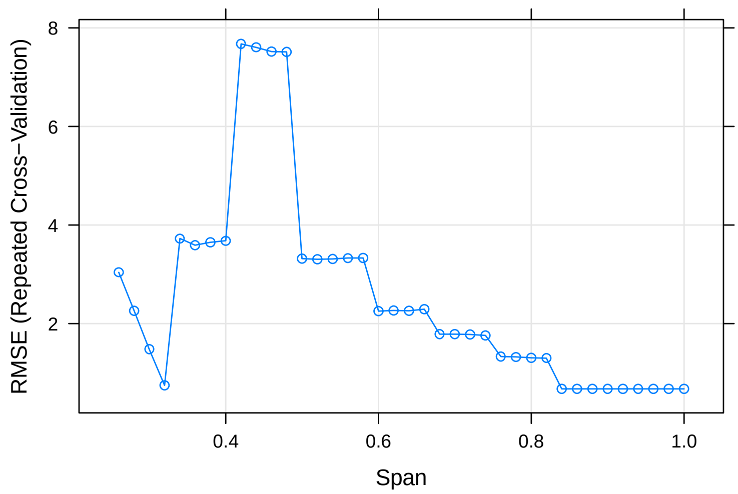


Figure 3. RMSE for gamLoess

2.1.4 Splines

A different smoothing can also be achieved using splines. Instead of using kernels, the independent variables are steadily transformed using splines before integrated in the GAM. The model is tuned upon the degrees of freedom parameter, which controls the degrees of freedom for the spline function (the more degrees of freedom, the higher the adaption to local structures). Two degrees of freedom lead to a fit with linear regression. Analogous to the gamLoess algorithm the results demonstrate that an adaption to local structures does not increase the model's accuracy. The best fit is achieved for df = 2, the linear model was already tested with the GLM (see Figure 4).

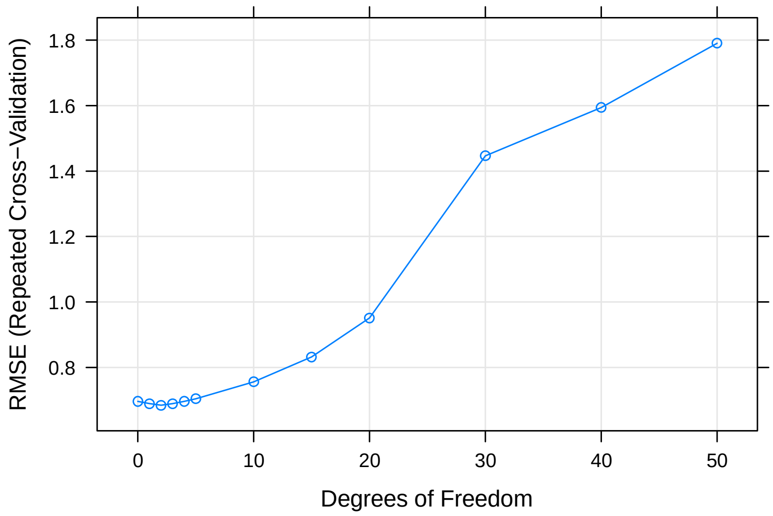


Figure 4. RMSE for gamSplines

Even though a small improvement using splines was expected and not achieved, the results are not astonishing as splines fit each independent variable within the GAM independently and are not capable of modeling interdependencies.

2.1.5 npreg

The most advanced kernel-smoothing algorithm applied in this study is computed upon the np-package in R. The npreg function computes a kernel for each independent variable and applies a local linear regression within the kernel. The optimal kernel parameters are independently data-driven optimized for each independent variable. Thereby a different bandwidth results for each of the independent variables. One of the most important advantages of this algorithm is that continuous as well as categorical, unordered variables (as present in this study) can be included in the regression (Racine, 2004). The algorithm is consequently capable of predicting with mixed datasets. It can either be computed with a Gaussian, an Epanechnikov or a linear kernel for continuous input data. Categorical data is calculated with an Aitchisonaitken or Liracine kernel. For this study, the categorical predictors (location, job and gender) were fitted upon Aitchisonaitken kernel only.

For each cross-validation, the kernel bandwidth for each input variable is computed via a Kullback-Leibler cross-validation or least-squares cross-validation, which is applied to compare algorithms upon RMSE in this study. In contrast, the Kullback-Leibler cross-validation compares different bandwidths upon the Akaike information criterion (AIC), which compares the goodness of fit with the model's complexity. As a result of bandwidth selection and parameter comparison, two nested cross-validations with correspondingly high computational costs have to be performed in order to test each bandwidth specification on several folds. The algorithm moreover uses either local-linear regression (ll) or the local-constant estimator (lc). The latter is an average smoother, similar to the k-nearest neighbor smoother, but contrarily computes different bandwidths and scale factors for each independent variable.

The results (see Figure 5a-b) show that the local-linear regression is more accurate than the local-constant estimator and reaches the GLM performance with the Epanechnikov kernel for least squares cross-validation (*RMSE = 0.682; RMSE; SD = 0.065*). The uniform kernel with local-linear regression Kullback-Leibler cross-validation does not reach sufficient accuracy (*RMSE > 5*), and is therefore excluded in the chart.

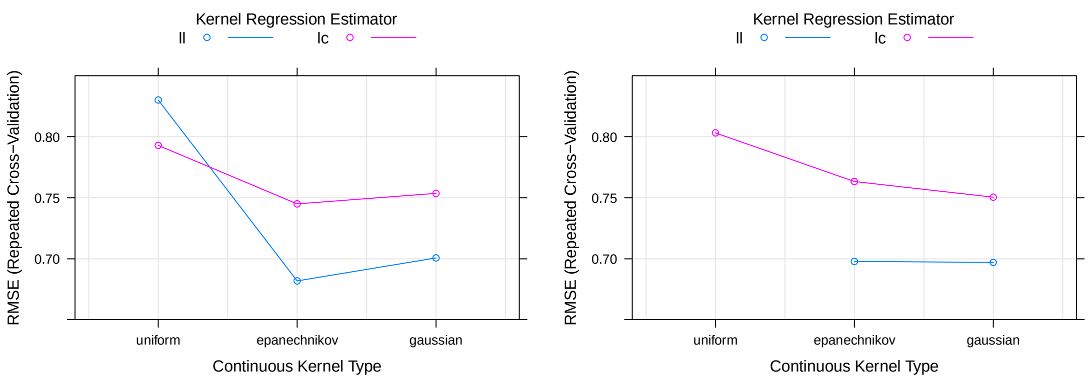


Figure 5a-b. RMSE for npreg with least-squares cross-validation (a) and Kullback-Leibler cross-validation (b)

Besides the models' accuracy, the variance between several cross-validation loops is an important aspect to evaluate the model's prediction capability. Reviewing the RMSE density plots finds that the Epanechnikov kernel provides the smallest variance between CV runs, followed by the Gaussian and then the linear kernel. For the local-constant estimator the variance is even smaller compared to the local-linear regression, but the latter performs better regarding RMSE mean (Figure 6).

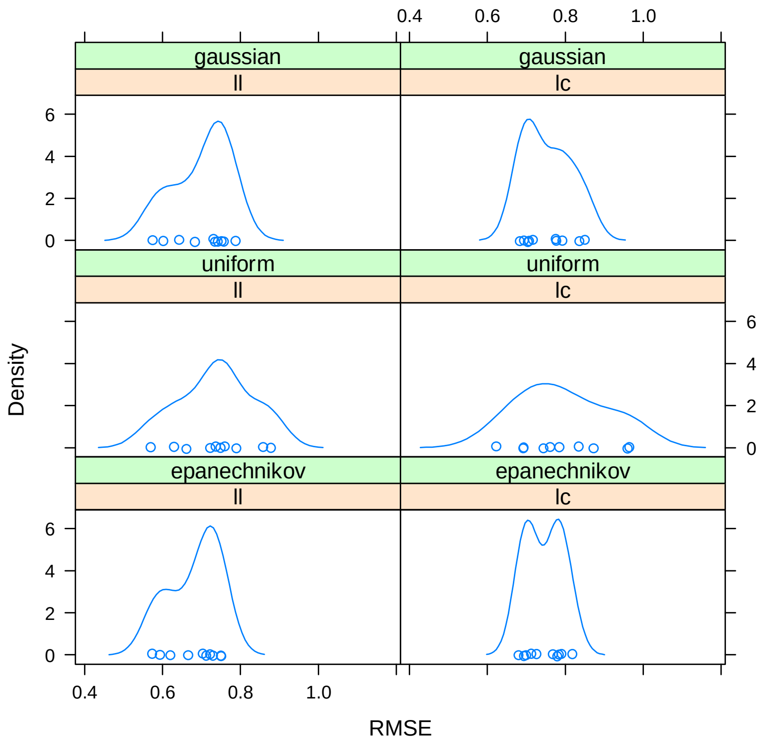


Figure 6. RMSE density plot for 10-fold cross-validation runs (kernel bandwidth selection upon least-squares cross-validation)

The algorithm has also been tested with higher kernel orders (kernel order = 2 and 4), but no accuracy gains were realized and consequently the following analyses apply secondary Epanechnikov kernels only.

Due to the variable bandwidth and scale estimations for the independent variables, npreg usually allows for an advanced analysis of the predictors' importance. Since the npreg algorithm does not predict the averaged well-being data more precisely than the GLM in this case, the variance importance just reflects the GLM predictor importance. However, the graphical representation in Figure 7 presents the partial, almost linear (kernel bandwidths >> n) regressions. The predictors were abbreviated to simplify the analysis.[46]

[46] Abbreviations: N - Neuroticism, E - Extroverted, A - Agreeableness, O - Optimism, C - Conscientious, M - Maximizer, F - Fairness, H - Health, Age - Age, L - Location, G - Gender, Edu - Education, J - Job.

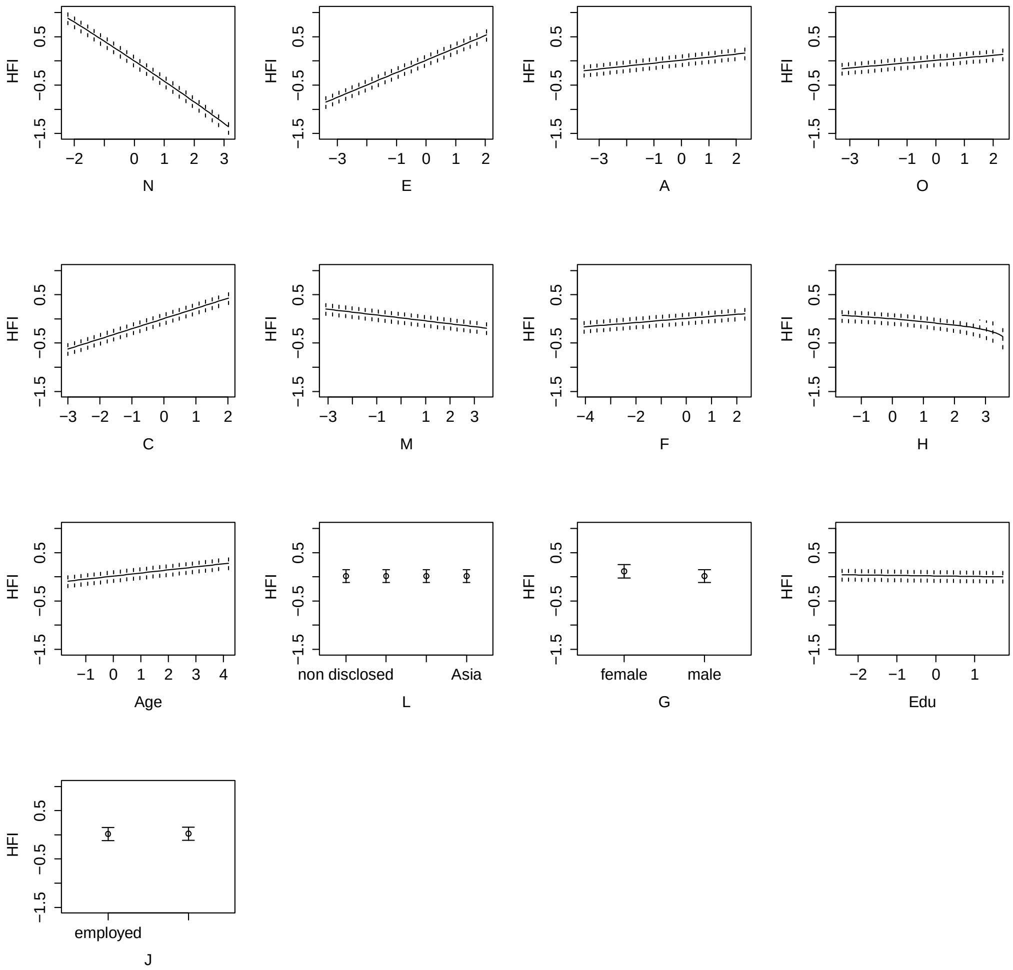


Figure 7. npreg predictors' partial regression influence

High dimensionality of the input data masks several non-linear linkages of certain independent variables. If less important independent variables are removed from the analysis, they come to light. Table 1 shows selected subsets of independent variables with reached performance measures. All calculations were conducted upon least-squares cross-validation with local linear regression within Epanechnikov kernels to fit the bandwidths and two times repeated 10-fold cross-validation to evaluate the performance. Due to the computational costs only a limited number of subsets could be tested.

Table 1. npreg accuracy for reduced input dimensionality (1)

Resampling results across tuning parameters:

subset	RMSE	Rsquared	RMSE SD	Rsquared SD
N E A O C H M F Age L G Edu J	0.702	0.51	0.0773	0.127
M H E A C N O F Age L G Edu J	0.762	0.513	0.285	0.131
N E A O C H M	0.691	0.522	0.0682	0.118
N E A O C	0.703	0.509	0.0695	0.128
O C E A N	0.701	0.513	0.0688	0.13
M N C E	0.692	0.522	0.0683	0.119
N E H M	0.708	0.499	0.0749	0.139
E M N H	0.709	0.498	0.0754	0.14
M E N H	0.708	0.499	0.0749	0.14
N E H	0.723	0.478	0.085	0.152
N E	0.728	0.471	0.0757	0.147
N H M	0.748	0.446	0.0713	0.133
N M H	0.744	0.452	0.0758	0.138
Age M N H E	0.704	0.505	0.0737	0.139
M H E A C N O F Age	0.7	0.512	0.0794	0.121
Age G J L	1	0.0367	0.0859	0.0373
N H	0.766	0.421	0.0829	0.138
N M	0.774	0.412	0.0584	0.11
N A	0.768	0.418	0.0638	0.114
N A E	0.724	0.477	0.0676	0.132

Tuning parameter 'regtype' was held constant at a value of ll
Tuning parameter 'nmulti' was held constant at a value of 10
Tuning parameter 'ckertype' was held constant
 at a value of epanechnikov
Tuning parameter 'ckerorder' was held constant at a value of 2
Tuning parameter 'bwmethod' was held constant at a value of cv.ls

It is found that certain subsamples of the input data achieve almost as good accuracy as the original model including all independent variables. This applies to RMSE as well as the RMSE SD. For example, the independent variables' subset including the big five personality traits, health and the maximizer vs. satisficer test achieved an error of RMSE = 0.691, which is only one per cent worse than the best full model fit. A graphical representation of the dependencies within this subsample fit is given in Figure 8**Fehler! Verweisquelle konnte nicht gefunden werden.**. The fact that subsamples of the independent variables reach similar accuracy leads to the conclusion that the correlation between the predictors has an influence when fitted locally.

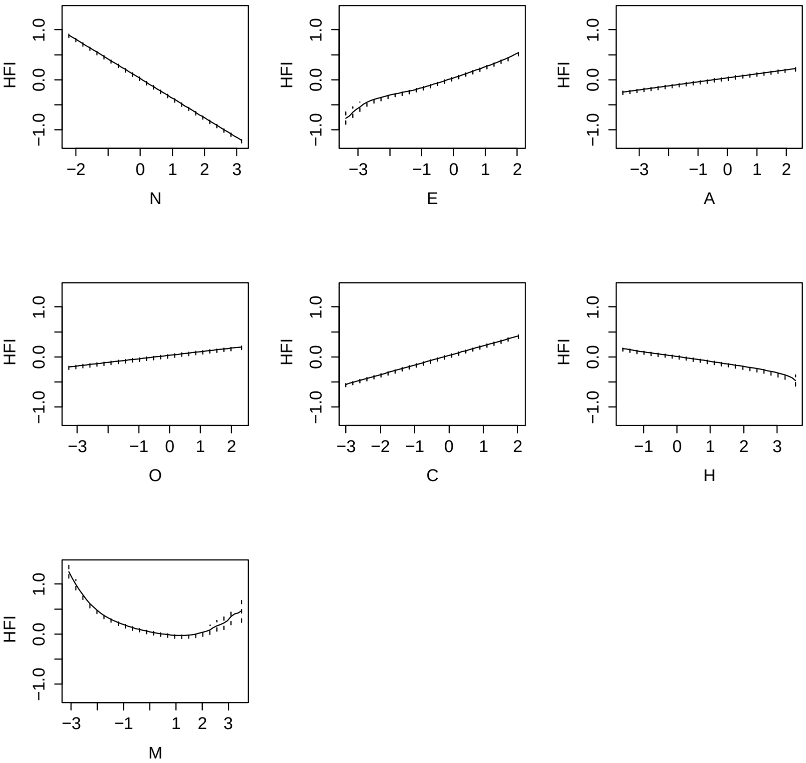


Figure 8. npreg predictors’ partial regression influence for reduced input dimensionality (1)

The maximizer-satisficer measure has been found to have a U-shaped partial influence in many subsets, even if the overall model fits almost linear (very large kernel bandwidth; see Figure 7). In contrast to the intuitive suggestion that maximizers have lower well-being than satisficers, maximizers seem to be happier than the average. This is even more supported, when age, as the predictor most correlated with the maximizer-satisficer variable is included in the model (Figure 9). Directly compared to the predictors conscientiousness and agreeableness, the maximizer-satisficer predictor explains less variance than conscientiousness (higher RMSE), but more than agreeableness (Table 2).

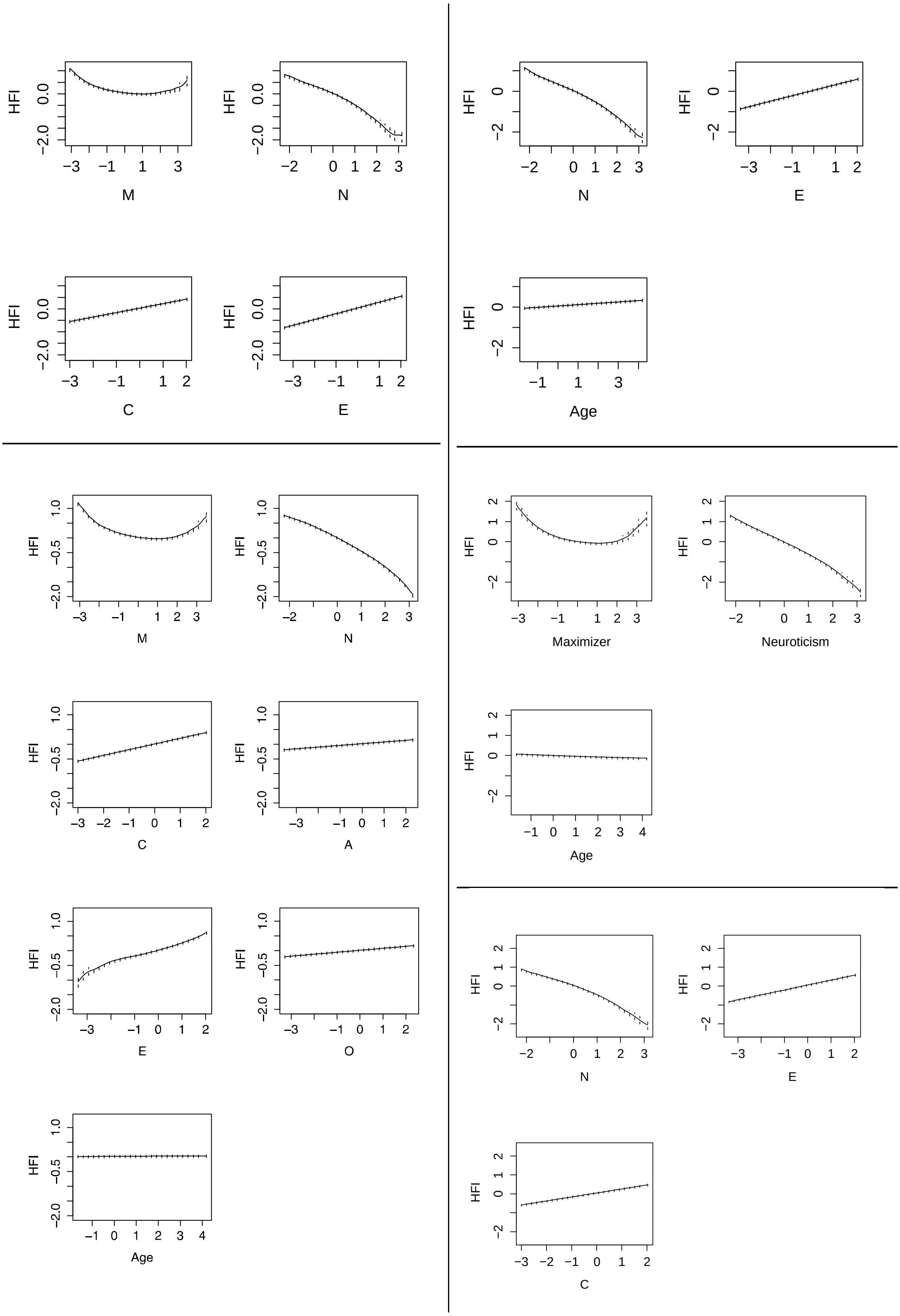


Figure 9. npreg predictors' partial regression influence for reduced input dimensionality (2)

Table 2. npreg predictors' partial regression influence for reduced input dimensionality (3)

subset	RMSE	Rsquared	RMSE SD	Rsquared SD
N E	0.735	0.459	0.0628	0.145
N E C	0.697	0.507	0.0745	0.135
N E M	0.715	0.491	0.0583	0.139
N E A	0.726	0.471	0.0596	0.131

The overall model shows a small positive linear influence of age, but those results are not obtained from long-time measurement and are consequently not corrected for influences by different cohorts. Moreover, the negative influence of a healthy lifestyle already identified by the GLM was confirmed by non-parametric regression. None of the calculated predictor's subsets showed a positive influence of a healthy lifestyle.

An interesting observation was made when the predictors were reordered. The algorithm results in different accuracies for different predictor orders which are stable during cross-validation. The algorithm calculates different bandwidths for different predictor orders.

2.2 Neural Network Algorithms

2.2.1 Stuttgart Neural Network Simulator

The neural networks applied in this study are implemented using the Stuttgart Neural Network Simulator (SNNS) package in R. In order to perform the same cross-validated analyses as for the before mentioned algorithms, a custom model was built to integrate a fully customizable version of the SNNS into the caret package.

The SNNS allows for a variety of different learning algorithms, of which *standard backpropagation (SBP)*, the most common NN learning algorithm, and *scaled conjugate gradient (SCG)* has also been applied. Both perform supervised learning for feed forward neural networks, but differ in the optimization routine. While SBP uses the first derivative of the goal function, SCG optimizes upon the second derivative, which is computational more expensive, but generally finds a better way to the (local) minimum. SCG is a combination of a conjugate gradient approach and ideas of the Levenberg-Marquardt algorithm. Regarding the different learning algorithms' performance and accuracy, no clear ranking persists in the literature so far. Consequently, comparable studies usually apply and compare several different learning algorithms in order to find algorithms fitting the data best.

Due to the characteristics of the neural computing the dependent and independent variables have been normalized to zero mean and SD one. The categorical variables (e.g. gender, age, education) were consequently transformed to numeric variables. The neural network has been constructed with one to five hidden layers and 20 to 1000 nodes on each layer. For standard backpropagation the parameters have been kept fix on a level for best accuracy and rather high computational costs, which is due to the small sample acceptable: the learning rate at a low level of 0.1 and the maximum output difference at zero.

The achieved accuracy with different learning algorithms is given in Figure 10. It is found that none of the tested network layouts and none of the applied learning algorithms reaches better performance than the GLM. The neural network with four hidden layers and 40 hidden nodes each performed best and reached a minimum RMSE of 0.765 for the SCG learning function and a RMSE of 0.763 for the standard backpropagation learning function. Both learning functions provide very similar results.

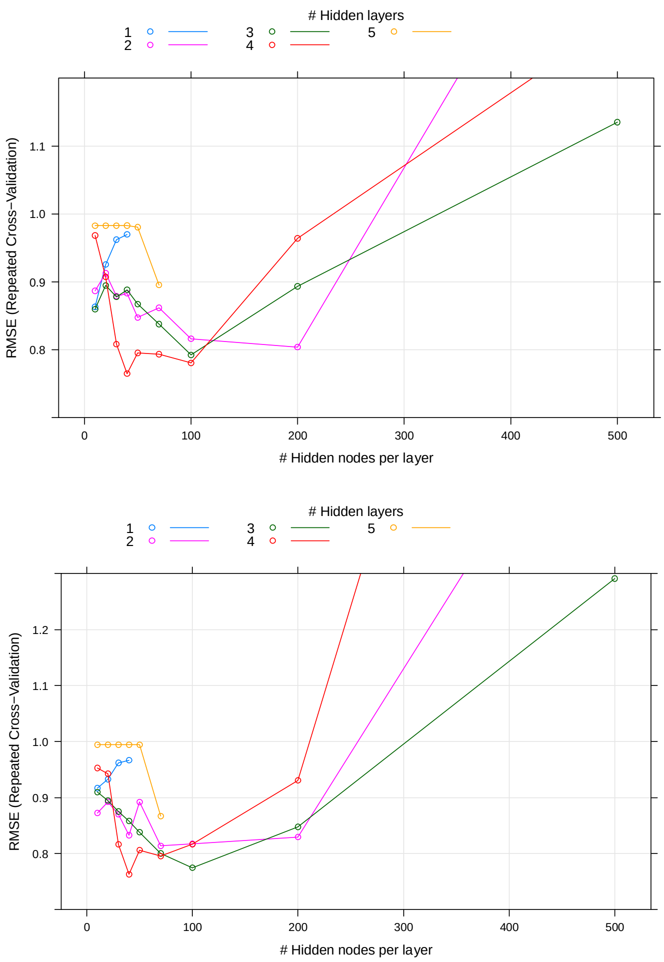


Figure 10. RMSE accuracy for feedforward neural network with SCG learning algorithm (a) and standard backpropagation learning algorithm (b) (learning rate = 0.1 and maximum difference = 0)

2.2.2 Extreme Learning Machine

Standard feedforward neural networks as implemented by SNNS generally face issues of slow learning speed (backpropagation) and customizable learning functions with a high number of crucial parameters to set. A new method fitting neural networks has therefore been developed:

Extreme learning Machines (ELM) fit single-hidden layer feedforward neural networks upon mathematical, non-iterative solving only. The input weights for each hidden note are randomly chosen and not adapted, so that training is omitted. Training is only applied to the weights for the output calculation, which is computationally less costly and can consequently magnititudes of order faster than conventional methods. By an increase of the number of hidden nodes with random inputs weights the ELM is theoretically as powerful as conventional neural networks and capable of approximating any continuous target functions.

The elmNN package in R allows for the training of ELMs with different activation functions (sigmoid function for standard neural networks). For this study five activation functions have been tested for the hidden and the output nodes: sigmoid (sig), slightly steeper tan-sigmoid (tansig), stepwise 0 / 1 function hard-limit (hardlim), stepwise -1 / 1 function symmetric hard-limit (hardlims) and a pure linear function (purelin). For a comparison of the activation functions with different numbers of hidden nodes see Figure 11. The pure linear activation function obviously explains the same variance as the GLM and leads once more to the best fitting model. All fitting was conducted upon 5 times repeated 10-fold cross validation.

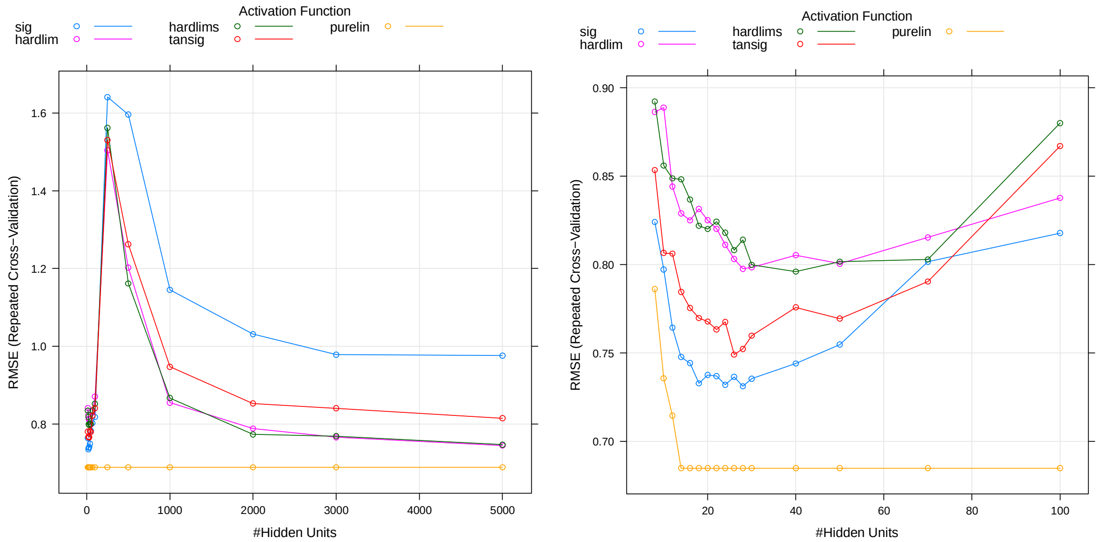


Figure 11. RMSE accuracy for extreme learning machine (ELM); right: zoom for small number of hidden nodes

Since the tansig activation function has, even for 5000 hidden nodes, been found to show decreasing RMSE with increased number of nodes, a single 5 times repeated 10-fold cross-validated analysis has been conducted for 12000 hidden nodes. However, it was still found that the sigmoid based activation functions do not outperform the GLM (Table 3).

Table 3. RMSE accuracy for extreme learning machine for 12,000 hidden nodes

```
Extreme Learning Machine

358 samples
 13 predictors

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)

Summary of sample sizes: 324, 322, 322, 322, 322, 322, ...

Resampling results across tuning parameters:

  actfun    RMSE   Rsquared  RMSE SD  Rsquared SD
  sig       0.957  0.283     0.103    0.124
  hardlim   0.724  0.472     0.0863   0.134
  hardlims  0.727  0.469     0.0857   0.137
  tansig    0.8    0.388     0.0786   0.121
  purelin   0.676  0.531     0.0792   0.136

Tuning parameter 'nhid' was held constant at a value of 12000
```

All tests have been conducted with 20 times-repeated 10-fold cross-validation. Since the hidden nodes input weights were randomly set, a sufficient number of repeated analyses have to be performed in order to achieve a valid accuracy result.

Due to the computational efficiency in combination with comparable accuracy, the ELM has also been applied to test for possible structures within each participant’s well-being trajectory. As already obtained from the GLM analysis no variance between the participants’ internal SD and internal regression coefficient (slope) of the linear trajectory smoothing could be explained (see Figure 12). All models upon the tested parameter sets result in higher RMSE than the samples SD (*RMSE > 1*).

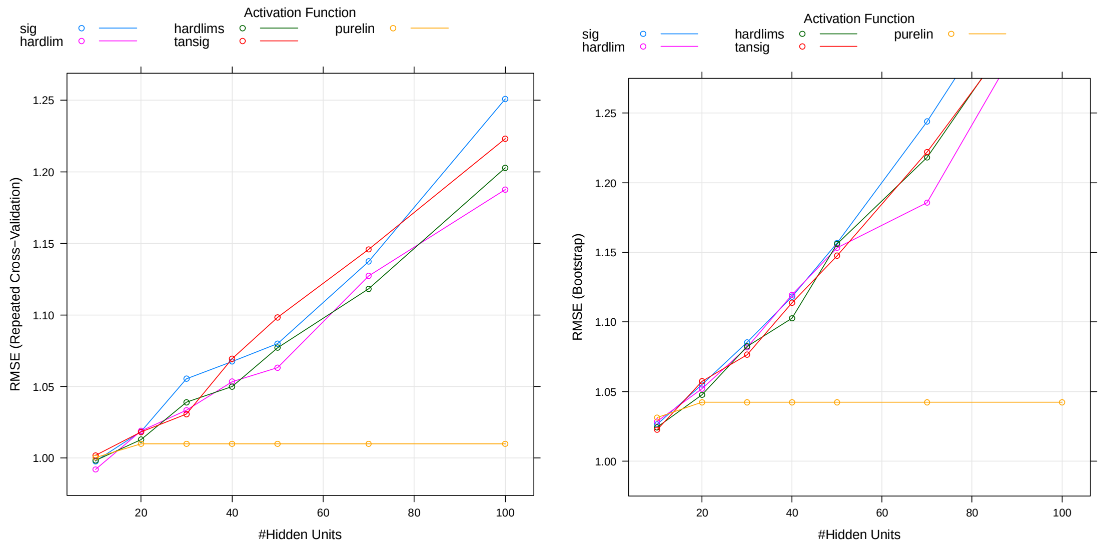


Figure 12. RMSE accuracy for ELM in trajectory prediction problem (left: SD as dep. var., right: reg. coefficient as dep. var.)

2.3 Feature Selection Algorithms

The following section does not aim for an accurate prediction of the independent variable. Instead, feature selection algorithms evaluate the importance of certain predictors for the output variable. The deployed kernel-smoothing algorithms indicate that certain independent variables within this study do not have an important influence on well-being. To evaluate this in detail, two different feature selection algorithms were applied.

2.3.1 Lasso and Elastic Net Regression

The lasso regression is a basic feature selection algorithm for generalized linear models (GLM). In comparison to algorithms using regularization the lasso algorithm limits the sum of coefficients (l_1 norm) to a constant and therefore results in coefficients being actually zero. The lasso regression is parameterized by the fraction of the full model coefficients' (l_1 norm), defining a maximum threshold for the sum of the current regression coefficients' (l_1 norm). A fraction of 1 consequently results in the full GLM, while a fraction of 0 forces all coefficients to zero. The algorithm is implemented using the lars and elasticnet package in R and 5 times repeated 10-fold cross-validated. Figure 13 outlines the lasso regression path and accuracy.

Sequence of LASSO moves:

	Neuroticism	Extroverted	Conscientious	Health	Maximizer	Optimisim	Agreeablness	Gender	Age	Job	Location	Fairness	Education
Var	1	2	5	8	6	4	3	11	9	13	10	7	12
Step	1	2	3	4	5	6	7	8	9	10	11	12	13

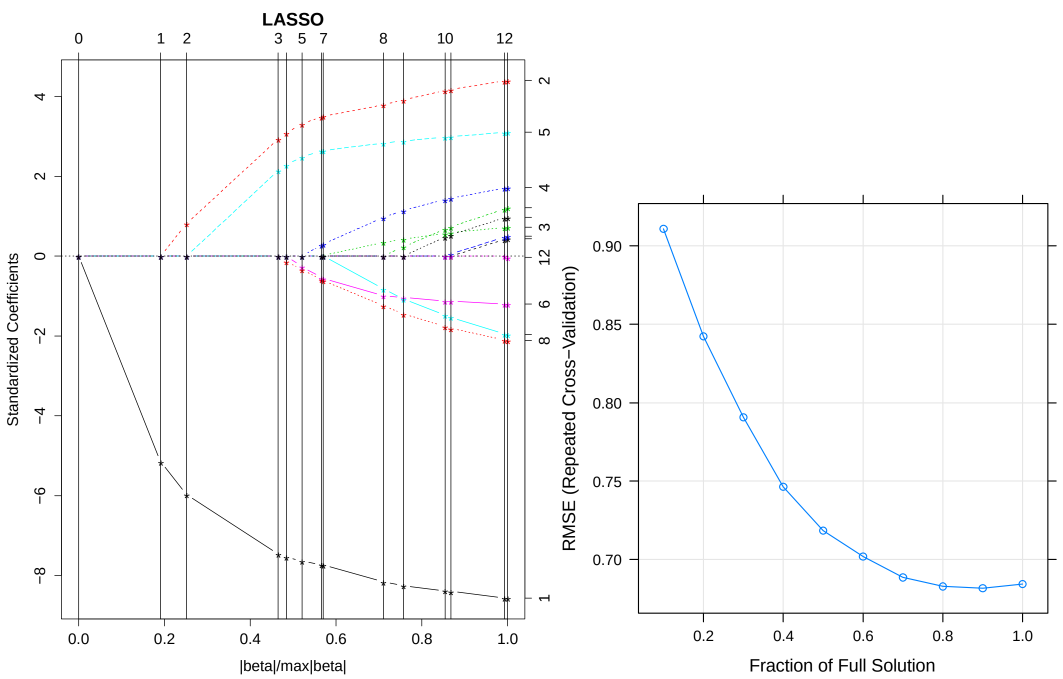


Figure 13. Lasso regression path (left) and RMSE accuracy (right)

As expected, the RMSE of the model approaches the GLM accuracy for the full solution. From the RMSE plot, a small improvement to the GLM can be observed, if the fraction is set to 0.9, so that fairness and education are not part of the model. It is concluded that these variables actually explain no structural variance in the linear model and hence overfit the data. The lasso path includes neuroticism as first, extroversion as second and conscientiousness as third variable.

Further developments of the lasso regression led to alternative norms for coefficient regularization. The Elastic Net Regression allows for continuous adjustment of the regularization norm including l_1 and l_2 norm by the parameter λ. However, for this study the elastic net regression including a parameterization for ridge regression did not provide an improvement in accuracy or feature selection.

2.3.2 Lazy Lasso Regression

The lazy lasso algorithm has been developed to combine kernel-smoothing with lasso regression. The combination allows fitting non-linear functions upon the locally most important independent variables only. Since the algorithm implements the lasso algorithm mentioned before, it actually zeroes unimportant regression coefficients by fitting the local lasso regression with the lars R package. However, the lazy lasso algorithm is not available as an R package yet, a simple version with a uniform kernel has been implemented.

Additionally, the algorithm is cross-validated using the caret package in order to test different parameter sets. The parameters include the bandwidth parameter t for the uniform k-nearest neighbor kernel (number of neighbors included) and a stopping parameter k, which defines the number of loops in a row to be calculated without performance improvements until the algorithm aborts. For each iteration the distances for the kernel calculation are parameter-wisely weighted with the regression coefficients from the previous iteration. The first iteration starts without weighting. This approach attaches more importance to relevant variables because distances by irrelevant predictors are neglected. In order to parameterize the distance adjustment, the calculation of δ_j is as follows:

$$\delta_{j=p*} \frac{|\beta_j|^{\,d}}{\sum_{j'=1}^{p}|\beta_{j'}|^{\,d}}$$

This allows for a scaling of the adjustment’s power by the distance adaption parameter d. For d = 1, δ is equal to the relative predictor weight; for d = 0, δ equals 1 for each predictor, so that no adjustment of the kernel to the predictor weight takes place.

As the algorithm performs feature selection upon the Lasso regression, a criteria to define the number of predictors included in the local linear regression is necessary. Upon the residual standard error for each step of the lars path Mallows’ C_p statistic is calculated. Predictors are included in the final model as long as C_p is larger than the total number of predictors multiplied by a bias factor, which is bias = 1 for the standard configuration, but may be parameterized. A larger bias factor results in a less complex model, a smaller bias factor includes more predictor variables.

Due to feature selection, the model's achieved accuracy is not comparable with the prediction models mentioned previously. However, the results from the parametric optimization can be gained from Figure 14. As expected, the kernel-smoothing demonstrates once more that the best model is achieved for large kernels approaching the generalized linear model. The stopping parameter k was tested for values k = 5 and k = 8 without noticeable differences, so that it is fixed to k = 5 for all further analysis.

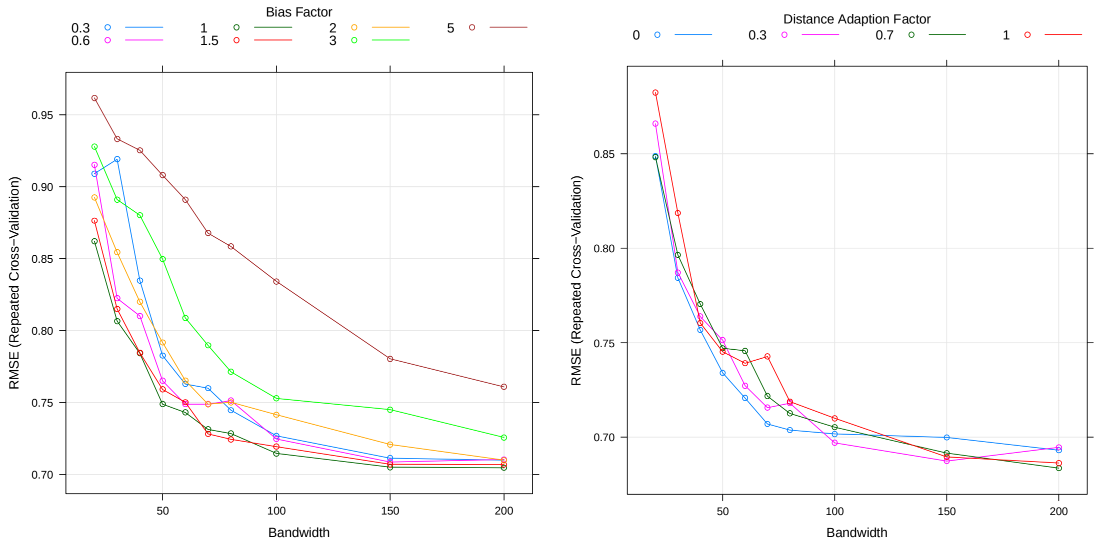


Figure 14. RMSE accuracy for lazy lasso regression (left: d = 1; right: bias = 1)

The bias factor was, as expected, found to reduce the number of predictors included in the local linear regressions and consequently reduces the accuracy when increased. Different from original expectations, the distance adaption factor d had a rather small influence on the model's accuracy. For medium-sized kernels (30 - 80 points) models with little distance scaling actually fitted the testing points better than the proposed distance scaling with d=1. Moreover, those models generally included fewer variables on average.

In order to evaluate the predictors' importance the final local regression coefficients for each testing point are saved and allow for later statistical analysis, for example counting the regressions with coefficients unequal to zero for each participant or sum the absolute regression coefficients by parameter. However, since the best performing model has a large kernel, those feature selection results are similar to the variance importance identified by the GLM. Hence, the assessment of the local predictor importance has been conducted on models with 30 to 80 points per kernel, even if those were not performing best in terms of accuracy. Figure 15a provides an overview of the predictor weights depending on the bias factor. It can be observed that neuroticism is the predominant predictor gaining even more importance, if the restriction is tightened (higher bias). Extroversion and conscientiousness were found to be the second most important predictors. However, their influence decreases, when the kernel size is shrunken and the prediction consequently based on fewer neighbors. This is different than expected, because a local analysis usually increases the relative importance of generally less important variables. Even for kernels with less than 30 points (< 10% of the sample size) neuroticism is the only important predictor. Extraordinarily increased weights for other predictors are not observed. However, the unrestricted model (bias = 0) for small kernels weights all predictors relatively equal with five to 15 per cent of the total predictor weight[47]. As seen in Figure 15b this includes an increased weight for the location variable. This has to be treated with caution, because the underlying sample is not representative in this regard. Moreover, the gender variable is comparably important in the unrestricted model with large kernel drops weight, when fitted locally.

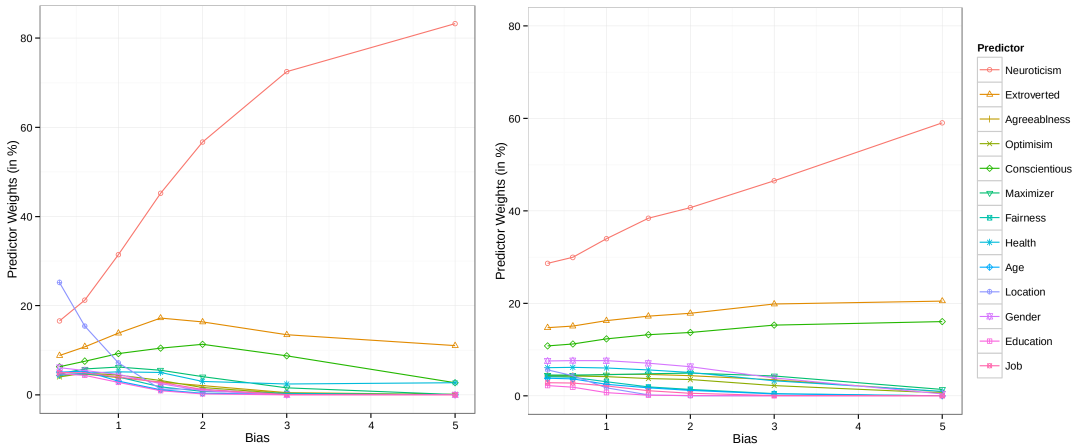


Figure 15a-b. Lazy lasso predictor weights (left: t \in [30,80], right: t \in [150,200])

Since the lasso regression actually zeros unimportant predictors when called with sufficient restriction via the bias variable, an analysis of the number of coefficients unequal to zero per predictor over all testing points is promising, too. Again, neuroticism, extroversion and conscientiousness stack out as the most often included predictors, followed by health and the maximizer-satisficer measure (Figure 23). When fitted locally with small kernel size, the differences between predictors are less distinct. For an average number of 2.5 predictors neuroticism is for example included in 40% of all local fitted regressions with small kernel (30 - 80 points) only, while included in over 65% of the regressions with larger kernels. Correspondingly, variables not important in larger kernels are included in local regressions with smaller kernels more often. Nevertheless, this is likely to result from over-fitting the data, since those small kernels result in significantly less cross-validated accuracy (Figure 16a-d).

[47] Note in this regard that the lars algorithm called for each local kernel environment individually shifts the training points to zero mean and variance one for each predictor.

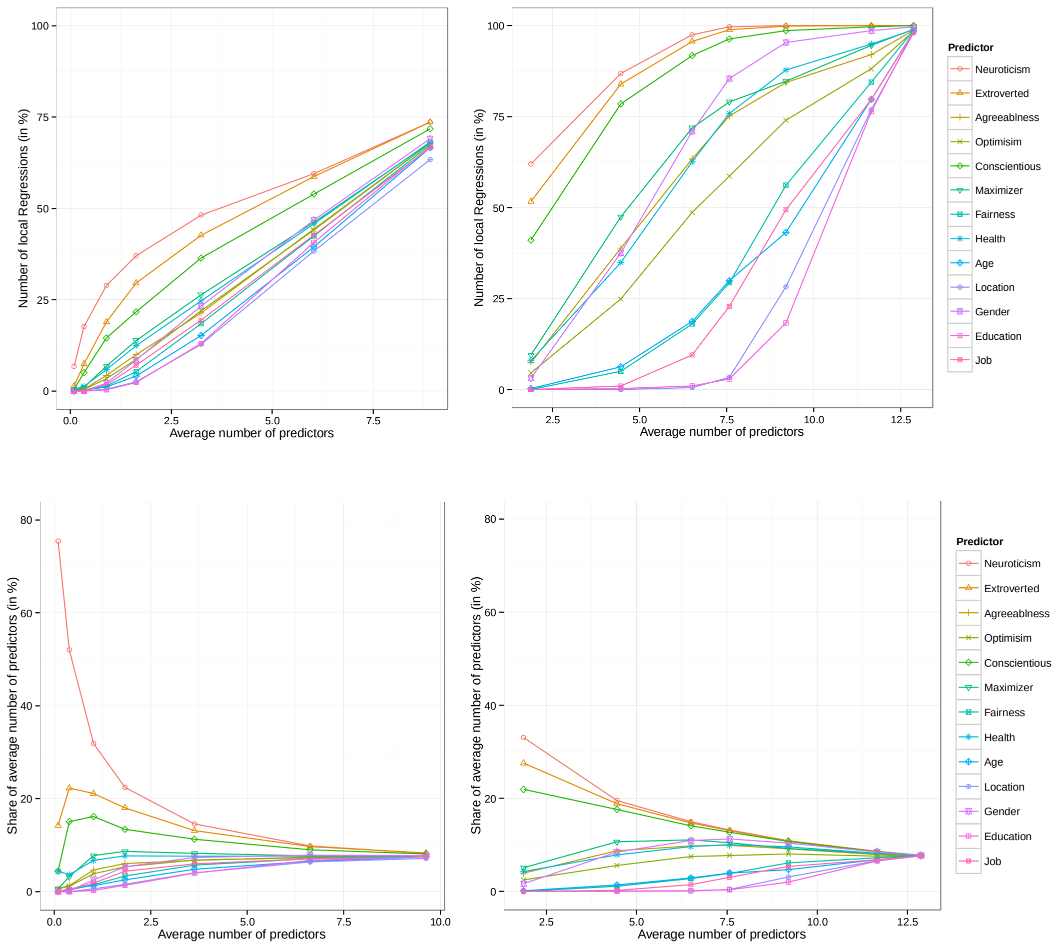


Figure 16a-d. Lazy lasso: percentage of local lasso regressions with predictor coefficient unequal to zero (left: t \in [30,80]; right: t \in [150,200]; top: measure relative to total number of regressions; bottom: measure relative to total number of regressions corrected with total number of predictors per regression)

In general, differences for the predictors' order concerning the frequency of coefficients unequal to zero is not observed with different kernel sizes. This once more supports that the high predictor weight of the location for small kernels is due to irregularities in the dataset. However, the variables can be clustered into three groups by importance, which are on the one hand fairly constant regarding the predictor weight and the frequency of coefficients unequal to zero and moreover correspond on the other hand with the finding from the npreg algorithm mentioned before (Table 4). Firstly, neuroticism, extroversion and conscientiousness explain by far most of the variance, neuroticism alone already around 40%, if fitted with non-parametric regression. Extroversion and conscientiousness add another ~ 10% of explained variance after controlling for neuroticism. The second group includes the maximizer-satisficer scale, health, optimism, agreeableness and gender. Especially for large kernels, the second group accounts for significantly more predictor weight than the remaining variables. Together with the first group, the variables explain approximately 47% of the variance between the averaged HFS per participant. The third group contains the remaining predictors fairness, education, job, location and age, which were found to have a rather small influence and explain very little variance after controlling for the groups one and two. Within the third group, age and fairness are the most relevant predictors. This division in three clusters is supported by the findings of the npreg algorithm and furthermore corresponds with the separation in the linear lasso regression on the whole dataset.

Table 4. Predictor importance by group. Note: Numbers in the second column indicate the difference between RMSE of model including the group as predictors and model including the more important groups only; analysis conducted with npreg algorithm.

	Predictors	RMSE contribution to full model	Variance explained as single predictor
Most important predictors (Group 1)	Neuroticism Extroversion Conscientiousness	0.40	41 % 22 % 15 %
Moderately important predictors (Group 2)	Maximizer Health Gender Agreeableness Optimism	0.04	8 – 12 %
Less important predictors (Group 3)	Age Fairness Job Education Location	0	0 – 8 %

While the lazy lasso algorithm is capable of effective feature selection and interpretation, it does not allow for an overall picture of a single predictor's influence as for example the npreg algorithm. The kernel-smoothing selects local environments around the predicted test points, but does not currently save the bandwidth information in order to compute the complete partial influence plot. Changes of local predictor importance along the predominant regression line of neuroticism could be subject to further research.

Since this study's sample is comparably small for the number of predictors included in the prediction models, an accuracy test for a reduced sample size is advised in order to test for possible accuracy advantages from larger datasets. This test has been conducted for the neural network model. The mentioned model was adapted to loop over different subsets of the sample and apply the cross-validated neural network algorithm on the subsets. Subsets including 50%

- 100% of the original dataset were tested. The neural network was built with the two best performing parameter sets identified before: three hidden layers with 100 nodes each and four layers with 40 nodes each. Results indicate that further increases of the sample size do not promise large accuracy improvements (Figure 17). The RMSE curve already flattens for training sets larger than 80% of the data available (362 points).

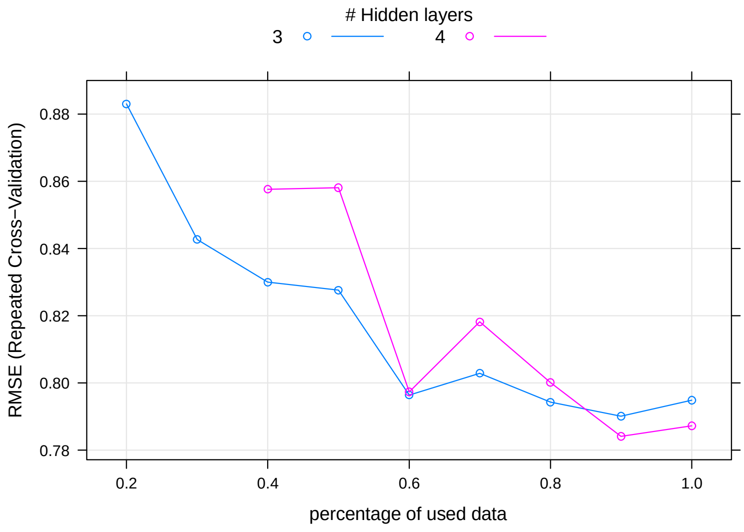


Figure 17. RMSE accuracy gains with increased number of training points for neural network

For further prove the same analysis has been conducted with the npreg algorithm. However, due to computational costs not the full 13-variable predictor set, but the seven most important predictors have been fitted. The results in Figure 18 support the implications previously mentioned. An extension of the dataset does not automatically lead to higher prediction results. Contrarily, the npreg algorithm almost achieves the maximum accuracy achieved in this study with 60% of the training data.

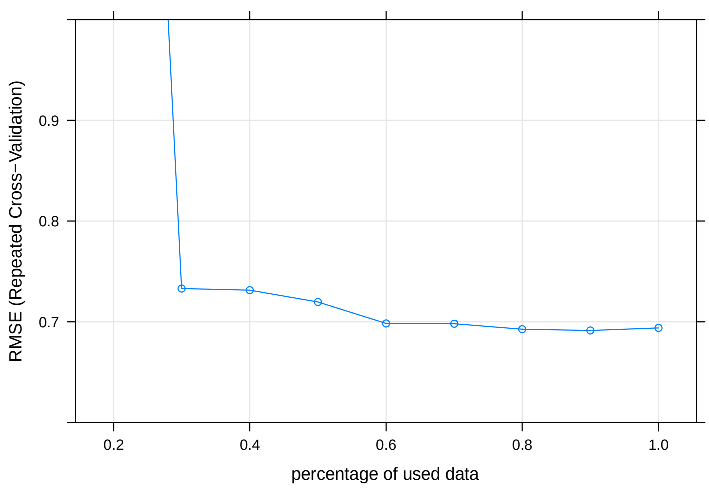


Figure 218. RMSE accuracy gains with increased number of training points for npreg

Appendix III Results of a Paired Sample t-test Considering Posts and Comments of Germany's Five Political Parties

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper			
Pair 1	CDUCSU_comments - CDUCSU_posts	.37750	.87798	.10975	.15819	.59681	3.440	63	.001
Pair 2	CDUCSU_comments - DIE_Linke_comments	-.02328	.20852	.02606	-.07537	.02880	-.893	63	.375
Pair 3	CDUCSU_comments - DIE_Linke_posts	.33047	.86925	.10866	.11334	.54760	3.041	63	.003
Pair 4	CDUCSU_comments - FDP_comments	.01953	.18108	.02263	-.02570	.06476	.863	63	.391
Pair 5	CDUCSU_comments - FDP_posts	.31187	.83760	.10470	.10265	.52110	2.979	63	.004
Pair 6	CDUCSU_comments - Grüne_comments	.04047	.15789	.01974	.00103	.07991	2.051	63	.044
Pair 7	CDUCSU_comments - Grüne_posts	.40281	.82997	.10375	.19549	.61013	3.883	63	.000
Pair 8	CDUCSU_comments - SPD_comments	-.02422	.17064	.02133	-.06684	.01840	-1.135	63	.260
Pair 9	CDUCSU_comments - SPD_posts	.32328	.79619	.09952	.12440	.52216	3.248	63	.002
Pair 10	CDUCSU_posts - DIE_Linke_comments	-.40078	.86726	.10841	-.61742	-.18415	-3.697	63	.000
Pair 11	CDUCSU_posts - DIE_Linke_posts	-.04703	.27204	.03400	-.11498	.02092	-1.383	63	.172

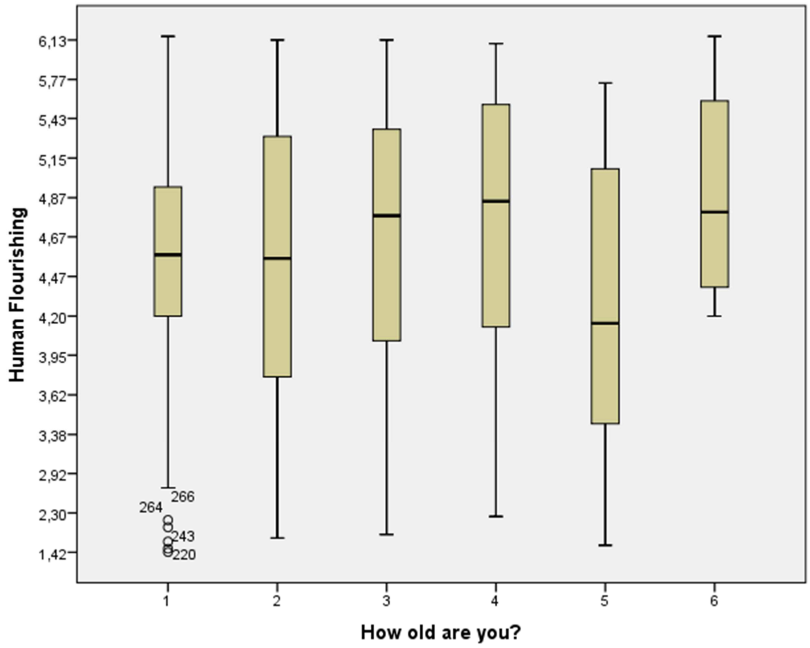
Pair 12	CDUCSU_posts - FDP_comments	-.35797	.85170	.10646	-.57072	-.14522	-3.362	63	.001
Pair 13	CDUCSU_posts - FDP_posts	-.06563	.29366	.03671	-.13898	.00773	-1.788	63	.079
Pair 14	CDUCSU_posts - Grüne_comments	-.33703	.82788	.10348	-.54383	-.13023	-3.257	63	.002
Pair 15	CDUCSU_posts - Grüne_posts	.02531	.25991	.03249	-.03961	.09024	.779	63	.439
Pair 16	CDUCSU_posts - SPD_comments	-.40172	.88207	.11026	-.62205	-.18139	-3.643	63	.001
Pair 17	CDUCSU_posts - SPD_posts	-.05422	.15282	.01910	-.09239	-.01604	-2.838	63	.006
Pair 18	DIE_Linke_comments - DIE_Linke_posts	.35375	.82152	.10269	.14854	.55896	3.445	63	.001
Pair 19	DIE_Linke_comments - FDP_comments	.04281	.13607	.01701	.00882	.07680	2.517	63	.014
Pair 20	DIE_Linke_comments - FDP_posts	.33516	.79225	.09903	.13726	.53306	3.384	63	.001
Pair 21	DIE_Linke_comments - Grüne_comments	.06375	.15537	.01942	.02494	.10256	3.282	63	.002
Pair 22	DIE_Linke_comments - Grüne_posts	.42609	.82469	.10309	.22009	.63209	4.133	63	.000
Pair 23	DIE_Linke_comments - SPD_comments	-.00094	.10574	.01322	-.02735	.02547	-.071	63	.944
Pair 24	DIE_Linke_comments - SPD_posts	.34656	.77837	.09730	.15213	.54099	3.562	63	.001
Pair 25	DIE_Linke_posts - FDP_comments	-.31094	.80137	.10017	-.51111	-.11076	-3.104	63	.003
Pair 26	DIE_Linke_posts - FDP_posts	-.01859	.17153	.02144	-.06144	.02425	-.867	63	.389

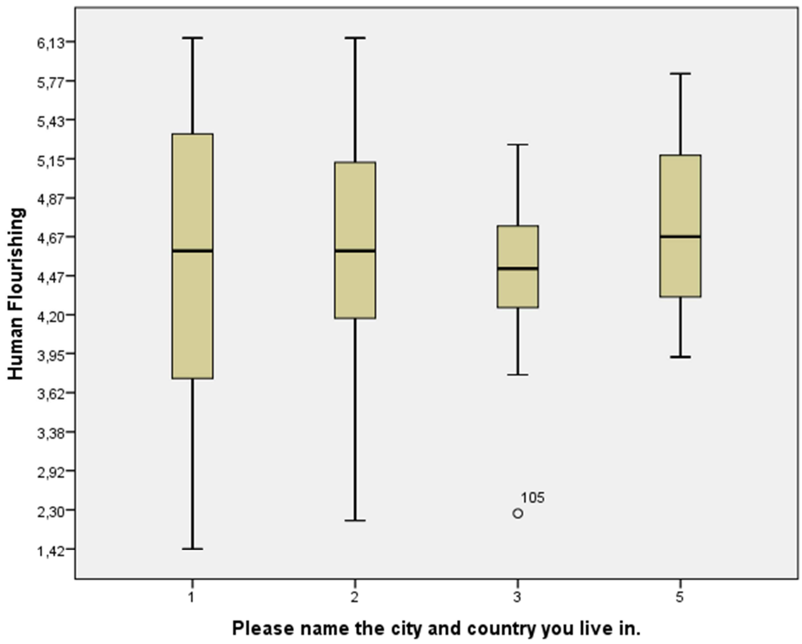
Pair 27	DIE_Linke_posts - Grüne_comments	-.29000	.79408	.09926	-.48836	-.09164	-2.922	63	.005
Pair 28	DIE_Linke_posts - Grüne_posts	.07234	.28742	.03593	.00055	.14414	2.014	63	.048
Pair 29	DIE_Linke_posts - SPD_comments	-.35469	.84619	.10577	-.56606	-.14332	-3.353	63	.001
Pair 30	DIE_Linke_posts - SPD_posts	-.00719	.19851	.02481	-.05677	.04240	-.290	63	.773
Pair 31	FDP_comments - FDP_posts	.29234	.77422	.09678	.09895	.48574	3.021	63	.004
Pair 32	FDP_comments - Grüne_comments	.02094	.09772	.01221	-.00347	.04535	1.714	63	.091
Pair 33	FDP_comments - Grüne_posts	.38328	.79445	.09931	.18483	.58173	3.860	63	.000
Pair 34	FDP_comments - SPD_comments	-.04375	.11730	.01466	-.07305	-.01445	-2.984	63	.004
Pair 35	FDP_comments - SPD_posts	.30375	.75652	.09456	.11478	.49272	3.212	63	.002
Pair 36	FDP_posts - Grüne_comments	-.27141	.77293	.09662	-.46448	-.07833	-2.809	63	.007
Pair 37	FDP_posts - Grüne_posts	.09094	.29145	.03643	.01813	.16374	2.496	63	.015
Pair 38	FDP_posts - SPD_comments	-.33609	.81996	.10249	-.54091	-.13127	-3.279	63	.002
Pair 39	FDP_posts - SPD_posts	.01141	.22669	.02834	-.04522	.06803	.403	63	.689
Pair 40	Grüne_comments - Grüne_posts	.36234	.76808	.09601	.17048	.55420	3.774	63	.000
Pair 41	Grüne_comments - SPD_comments	-.06469	.12972	.01622	-.09709	-.03228	-3.989	63	.000
Pair 42	Grüne_comments - SPD_posts	.28281	.73739	.09217	.09862	.46701	3.068	63	.003
Pair 43	Grüne_posts - SPD_comments	-.42703	.84078	.10510	-.63705	-.21701	-4.063	63	.000

Pair 44	Grüne_posts - SPD_posts	-.07953	.24361	.03045	-.14038	-.01868	-2.612	63	.011
Pair 45	SPD_comments - SPD_posts	.34750	.79106	.09888	.14990	.54510	3.514	63	.001

Appendix IV Descriptive Aspects of the AMT Survey Population Considering Mean HFS

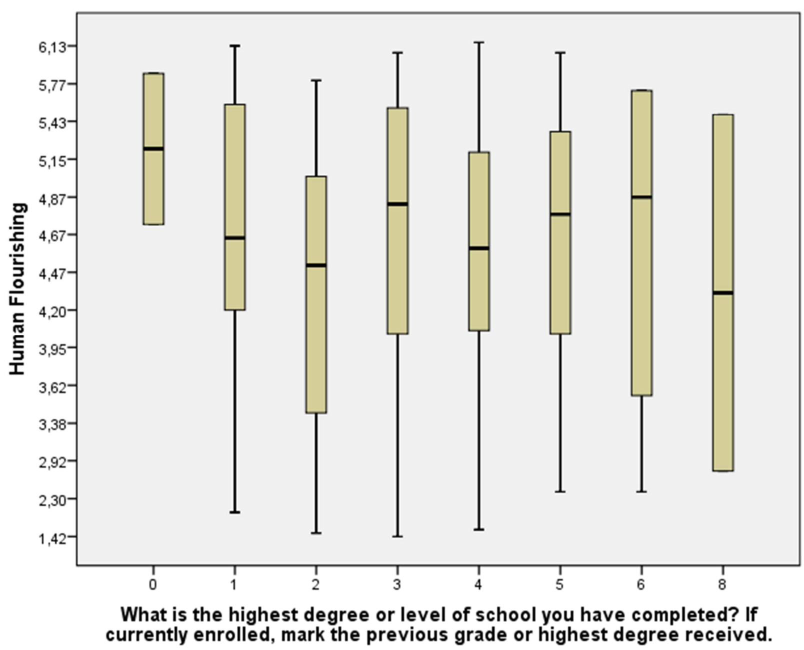
The below boxplots indicate some of the descriptive aspects of the AMT survey population. Under consideration are Human Flourishing Scores, age, gender, location, employment status, and highest education level.





Human Flourishing
6,13
5,77
5,43
5,15
4,87
4,67
4,47
4,20
3,95
3,62
3,38
2,92
2,30
1,42
105
1
2
3
5
Please name the city and country you live in.

Human Flourishing
6,13
5,77
5,43
5,15
4,87
4,67
4,47
4,20
3,95
3,62
3,38
2,92
2,30
1,42
0
1
Are you currently employed?



Human Flourishing
6,13
5,77
5,43
5,15
4,87
4,67
4,47
4,20
3,95
3,62
3,38
2,92
2,30
1,42
0
1
2
3
4
5
6
8
What is the highest degree or level of school you have completed? If currently enrolled, mark the previous grade or highest degree received.

Appendix V List of KIT Facebook Pages and their Organization into Subgroups

	Address	Page Name
KIT allgemein	http://www.facebook.com/pages/Karlsruher-Institut-f%C3%BCr-Technologie-KIT/107624245965021	(KIT)
	https://www.facebook.com/UniKarlsruhe?rf=112388085446516	(Uni Karlsruhe)
	http://www.facebook.com/pages/House-of-Competence-HoC/359972890600	(KIT HoC)
	https://www.facebook.com/Studipilot	Studierendenwerk Karlsruhe AöR
	https://www.facebook.com/KITStudyVisuallyImpaired	Study Centre for the Visually Imparied Students
	https://www.facebook.com/erasmus.ka	(Erasmus Karlsruhe)
Rund um die Bibliothek	http://www.facebook.com/pages/KIT-Bibliothek/155989387749416	(KIT Bibliothek)
	http://www.facebook.com/pages/Ohrst%C3%B6psel-am-KIT/281204658625762	(Ohrst”psel am KIT)
	https://www.facebook.com/pages/KIT-Dreht%C3%BCr/437740246353305?fref=pb&hc_location=profile_browser	KIT Drehtür
	https://www.facebook.com/FundstuckeAusDerKITBibliothek/	Fundstücke aus der Bibliothek
Fachschaften	https://www.facebook.com/FachschaftWiWi	(Fachschaft Wirtschaftswissenschaft)
	http://www.facebook.com/pages/Fachschaft-Architektur-KIT/121823821230771	(Fachschaft Architektur)
	http://www.facebook.com/pages/Fachschaft-MaschinenbauChemieingenieurwesen-am-KIT/111583662190017	(Fachschaft Maschienenbau/ Chemieingenieurwesen)
	http://www.facebook.com/pages/Fachschaft-Sport-KIT/235706879823177	(Fachschaft Sport KIT)
	https://www.facebook.com/fsmi.kit	(Fachschaft Mathe/ Unfo KIT)
	https://www.facebook.com/fachschaftchembio	(Fachschaft Chemie/ Biologie KIT)
	https://www.facebook.com/GeistSoz	(Fachschaft GeistSoz)
	https://www.facebook.com/pages/Fachschaft-Bau/191020064257178	(Fachschaft Bau)
	https://www.facebook.com/pages/Fachschaft-Physik-an-	(Fachschaft Physik)

	der-Uni-Karlsruhe/154199824745188	
Hochschulgruppen	https://www.facebook.com/AFK.KA	(Hochschulgruppe Kino KIT/ Akademischer Filmkreis)
	https://www.facebook.com/debattekarlsruhe	(Hochschulgruppe Debatte Karlsruhe)
	https://www.facebook.com/Amnesty.Karlsruhe	(Hochschulgruppe Amnesty International)
	http://www.facebook.com/pages/Juso-Hochschulgruppe-Karlsruhe/276740170730?ref=stream	(Hochschulgruppe JuSo)
	http://www.facebook.com/pages/LEAN-Hochschulgruppe-am-KIT/136142666439378	(KIT Hochschulgruppe LEAN)
	http://www.facebook.com/pages/KIT-Hochschulgruppe-College-MV/284167611615533	(KIT Hochschulgruppe College MV)
	https://www.facebook.com/akaflieg.karlsruhe	(Hochschulgruppe Akademische Fliegergruppe)
	https://www.facebook.com/kit.international	(International Affairs/ Internationals)
	https://www.facebook.com/VWIESTIEM.KARLSRUHE?fref=pb&hc_location=profile_browser	VWI ESTIEM Karlsruhe
	https://www.facebook.com/abgedrehtKarlsruhe?fref=pb&hc_location=profile_browser	Abgedreht - Die Filmgruppe am KIT
	https://www.facebook.com/pages/KAMUN-Karlsruhe-Model-United-Nations/459879100709978?fref=pb&hc_location=profile_browser	KAMUN- Karlsruhe Model United Nations
	https://www.facebook.com/AIESEC.Karlsruhe?fref=pb&hc_location=profile_browser	AISESEC Karlsruhe
	https://www.facebook.com/kit.enactus?fref=pb&hc_location=profile_browser	Enactus KIT
	https://www.facebook.com/ewb.karlsruhe?fref=pb&hc_location=profile_browser	Engineers Without Borders - Karlsruhe Institute of Technology e.V.
	https://www.facebook.com/pages/fuks/89516690661?fref=pb&hc_location=profile_browser	fuks
	https://www.facebook.com/crashkursefuks	Crashkurse fuks
	https://www.facebook.com/bikev	Börseninitiative e.v.
	https://www.facebook.com/brainreset.kit/	Ophasen-Gruppe Chemiker&Biologen
	https://www.facebook.com/pages/Studenten-für-Kinder-Karlsruhe-eV-SfKa/	Studenten für Kinder e.V.

	https://www.facebook.com/kamaroengineering/	Kamaro Engineering
	https://www.facebook.com/studentec/	Studentec
	https://www.facebook.com/deltaKarlsruhe	delta
	https://www.facebook.com/group54ka	Group 54
	https://www.facebook.com/RISK.KIT/	Risiko Initiative Stochastik Karlsruhe e.V.
	https://www.facebook.com/aegeeka	AEGEE - European Students' Forum
	https://www.facebook.com/EWBIndiraGandhi	Sonne für ein Kinderheim-Indien HSG
	https://www.facebook.com/pages/Global-Marshall-Plan-Hochschulgruppe-Karlsruhe/	Global Marshall Plan HSG
	https://www.facebook.com/SchmitzKatzeImpro	Schmitz' Katze Improtheater
	https://www.facebook.com/kit.enactus.ziczac/	ZICzac - Zukunft, Integration, Chance - Enactus
	https://www.facebook.com/WollWerkKA	Wollwerk
	https://www.facebook.com/mercygroup/	Mercy Group - Ehrenamtliche HSG
	https://www.facebook.com/Sprechreizkit/	Sprechreiz - Enactus
	https://www.facebook.com/TheaBib/	TheaBib - Enactus
	https://www.facebook.com/pages/CreatING/	CreatING
	https://www.facebook.com/AkademischerVereinKyrillundMethod/	Akademischer Verein "Kyrill und Method"
	https://www.facebook.com/HayekClubKarlsruhe	Hayek Club HSG
	https://www.facebook.com/iaeste.germany.karlsruhe/	IAESTE LC Karlsruhe HSG
	https://www.facebook.com/OpticsStudentsKarlsruhe	OSKar - Optics Students Karlsruhe e.V. HSG
	https://www.facebook.com/renewable.energy.challenge	reech - renewable energy challenge HSG
	https://www.facebook.com/KITcarTeam	KITcar HSG
	https://www.facebook.com/KaRaceIng/info	(KaRaceIng)
	https://www.facebook.com/kine.Karlsruhe	Karlsruher Initiative zur Nachhaltigen Energiewirtschaft
	https://www.facebook.com/msv.kit/	Muslimischer Studentenverein Karlsruhe e.V.
Uni Sport/ Sportgruppen	https://www.facebook.com/KITSportClub	(KIT Sport Club)

	https://www.facebook.com/KITSCGEQUOS	(KIT SC Gequos)
	https://www.facebook.com/Waterpolo.KIT	(KIT Waterpolo)
	https://www.facebook.com/KitScHandball	(KIT SC Handball)
	https://www.facebook.com/uniliga.karlsruhe	(Uniliga Karlsruhe)
	https://www.facebook.com/heimspiel.am.KIT	(Heimspiel; Kneipe am KIT)
	https://www.facebook.com/hochschulrudern.karlsruhe?fref=pb&hc_location=profile_browser	Hochschulrudern Karlsruhe
	https://www.facebook.com/tourEucor/info?tab=page_info	TourEucor
	https://www.facebook.com/KitScFussball?fref=pb&hc_location=profile_browser	KIT SC
	https://www.facebook.com/pages/Sportfreunde-Oettinger/	Sportfreunde Öttinger
	https://www.facebook.com/KITSCEngineers	KIT SC Engineers
	https://www.facebook.com/KitScFussball	KIT SC Fußball
	https://www.facebook.com/KarlsruheStorm	Karlsruhe Storm Lacrosse
	https://www.facebook.com/kit.biergier	KIT Biergier Sportmannschaft
	https://www.facebook.com/pages/FoSS-SportsCamp/317569028341621	FoSS-SportsCamp
Institute/ Fachbereiche	https://www.facebook.com/KITInformatik	(Informatiker)
	http://www.facebook.com/pages/IfSS-Institut-f%C3%BCr-Sport-und-Sportwissenschaft-KIT/242380065791821	(KIT Institut f r Sport und Sportwissenschaften)
	https://www.facebook.com/KITInfobau	(KIT Fakult„t f r Informatik/ Infobau)
	https://www.facebook.com/pages/Institut-f%C3%BCr-Meteorologie-und-Klimaforschung-Forschungsbereich-Troposph%C3%A4re/1425205657754671	Institut für Meteorologie und Klimaforschung, Forschungsbereich Troposphäre
	https://www.facebook.com/pages/S%C3%BCddeutsches-Klimab%C3%BCro-am-Karlsruher-Institut-f%C3%BCr-Technologie/209452392507596?fref=pb&hc_location=profile_browser	Süddeutsches Klimabüro am Karlsruher Institut für Technologie
	https://www.facebook.com/KarlsruheServiceResearchInstitute?fref=pb&hc_location=profile_browser	KSRI
	https://www.facebook.com/pages/Karlsruhe-School-of-Optics-and-Photonics-KSOP-KIT/101876529856809?fref=pb&hc_location=profile_browser	Karlsruhe School of Optics and Photonics KSOP (KIT)
	https://www.facebook.com/regionalwissenschaft	Institut für Regionalwissenschaft

	https://www.facebook.com/KCETA.KSETA/	KCETA - KIT Center Elementary Particle and Astroparticle Physics
	https://www.facebook.com/InstitutITAS	Institute for Technology Assessment and Systems Analysis
	https://www.facebook.com/pages/Hector-School-of-Engineering-and-Management/	Hector School of Engineering and Management
	https://www.facebook.com/MICMoR.ResearchSchool/	MICMoR - Helmholtz Research School
	https://www.facebook.com/pages/Carl-Benz-School-of-Engineering/102884716417714	Carl Benz School of Engineering
	https://www.facebook.com/heika.research/	Heidelberg Karlsruhe Research Partnership
	https://www.facebook.com/DidaktikderMathematikKIT	(Didaktik f r Mathematik am KIT)
	https://www.facebook.com/ZAKKarlsruhe?fref=pb&hc_location=profile_browser	ZAK | Zentrum für Angewandte Kulturwissenschaft und Studium Generale
	https://www.facebook.com/ZentrumfuerMedialesLernen?fref=pb&hc_location=profile_browser	Zentrum für mediales Lernen
	https://www.facebook.com/WMKstudium	Wissenschaft Medien Kommunikation
	https://www.facebook.com/foruminwi?fref=ts	Forum INWI
Innovation/ Entrepreneurs/ Entwicklung	https://www.facebook.com/KITInnovation	(KIT Innovation)
	https://www.facebook.com/CIEKIT	(CIE (Center f r Innovation und Entrepreneurs))
	https://www.facebook.com/Pioniergarage	(Pioniergarage/ Entrepreneurs KIT)
Hochschulpolitik	https://www.facebook.com/UStA.KA	(Usta KIT)
	https://www.facebook.com/AKVS.KIT	(Arbeitskreis Verfasste Studierendenschaft KIT)
	https://www.facebook.com/fips.am.kit	fips am KIT
	https://www.facebook.com/tugendfuror	Tugendfuriös - Queerfeministischer Lesekreis

	https://www.facebook.com/rosalistekarlsruhe/	Rosa Liste Karlsruhe
	https://www.facebook.com/gahgkarlsruhe	GAHG: grün-alternative HSG Karlsruhe
	https://www.facebook.com/lhg.karlsruhe	Liberale Hochschulgruppe Karlsruhe
	https://www.facebook.com/galkarlsruhe	GAL - Grüne Alternative Liste am KIT
	https://www.facebook.com/AlternativeListe	Alternative Liste Karlsruhe
	https://www.facebook.com/RCDSKarlsruhe	Ring Christlich Demokratischer Studenten Karlsruhe
	https://www.facebook.com/Semesterzeiten	Für internationale Semesterzeiten am KIT
	https://www.facebook.com/AStA.KIT	(Allgemeiner Studierendenausschuss am KIT)
Karriere/ Berufseinstieg	https://www.facebook.com/KIT.CareerService?fref=ts	(KIT Career Service)
	https://www.facebook.com/pages/Zentrum-f%C3%BCr-Information-und-Beratung-zib-am-KIT/172511296106594	Zentrum für Information-und-Beratung-zib-am-KIT
	https://www.facebook.com/R2Bstudent	r2b-student
	http://www.facebook.com/pages/Personalentwicklung-am-KIT/146718152064171	(Personalentwicklung am KIT)
	https://www.facebook.com/ctjka?fref=pb&hc_location=profile_browser	catch-the-job
Arbeitskreise	https://www.facebook.com/talKITKarlsruhe	(talKIT; Wirtschafts- und Technologieforum am KIT))
	https://www.facebook.com/unitheater	(Theater Universit„t)
	https://www.facebook.com/ustaunifest?sk=wall&filter=1	(usta Unifest)
	https://www.facebook.com/SC2KIT	(KIT Starcraft 2 Tournament)
	https://www.facebook.com/KITalumni	(KIT Alumni)
	https://www.facebook.com/pages/Radio-KIT/187986998001375?fref=pb&hc_location=profile_browser	Radio KIT
	https://www.facebook.com/KarlsruherTransfer	Karlsruher Transfer
	https://www.facebook.com/LeoClubKarlsruhe	Leo Club

	https://www.facebook.com/Lehramt.at.KIT	Lehramt am KIT
	https://www.facebook.com/pages/Förderverein-der-Studierendenschaft-des-KIT/227038090726686	Förderverein der Studierendenschaft
	https://www.facebook.com/Vorlesungsverzeichnis	Vorlesungsverzeichnis
	https://www.facebook.com/pages/KIT-Interkulturell-Arbeit-und-Wirtschaft/	KIT Interkulturell
	https://www.facebook.com/pages/KIT-Doktorandeninitiative/	KIT Doktorandeninitiative
	https://www.facebook.com/startcampKA/	Startcamp KA
	https://www.facebook.com/TAjournal	Technikfolgenabschätzung und Praxis
	https://www.facebook.com/iMensaKarlsruhe	Mensa App
	https://www.facebook.com/pages/DKMS-Typisierungstag-am-KIT-Studenten-gegen-Blutkrebs/	DKMS-Typisierungstag-am-KIT-Studenten-gegen-Blutkrebs
	https://www.facebook.com/KEULE2012/	Keule 2012
	https://www.facebook.com/FFIKIT/	Freundeskreis für Informatik am KIT
	https://www.facebook.com/businessmasters/	Business Masters - International Case Studies
	https://www.facebook.com/InsideScienceKIT/	Inside Science Magazion
Musik	https://www.facebook.com/KITBigBand	KIT Big Band
	https://www.facebook.com/pages/KIT-Konzertchor/	KIT Konzertchor
Social	https://www.facebook.com/pages/Spotted-KIT/	Spotted KIT
	https://www.facebook.com/KIT.Spotted	Spotted KIT
	https://www.facebook.com/akkballkarlsruhe	AKK Ball
	https://www.facebook.com/pages/Karlsruher-Gespräche-2011/	Karlsruher Gespräche 2011
	https://www.facebook.com/pages/Verspottet-KIT/	Verspotted KIT
	https://www.facebook.com/nightline.karlsruhe/	Nightline Karlsruhe
	https://www.facebook.com/unifest.karlsruhe	(Unifest Karlsruhe)
	https://www.facebook.com/IslamMeetsKIT?fref=pb&hc_location=profile_browser	Islam meets KIT
	https://www.facebook.com/akk77	AKK

Appendix VI Results of the Nearest Neighbors Analysis for the KIT Facebook Network, *k=5*

KIT Group	k=1	k=2	k=3	k=4	k=5
Arbeitskreise comments	11.045 Hochschulegruppen comments	11.045 Fachschaften comments	11.045 KIT allegemein posts	11.136 Hochschulpolitik comments	11.136 Social comments
Fachschaften comments	10.954 Social comments	11.045 Innovation, Entrepreneurs, Entwickung posts	11.045 Uni Sports posts	11.045 Fachschaften posts	11.045 Hochschulgruppen posts
Hochschulgruppen comments	11.045 Social comments	11.045 Innovation, Entrepreneurs, Entwickung comments	11.045 Arbeitskriese comments	11.045 Fachschaften posts	11.045 Arbeitskreise posts
Hochschulpolitik comments	11.045 Social posts	11.136 Arbeitskreise comments	11.136 Rund um die Bibliothek comments	11.136 Institute, Fachbereiche posts	11.225 Uni Sports comments
Innovation, Entrepreneurs, Entwicklung comments	10.770 Social posts	10.863 Fachschaften posts	10.954 Uni Sports posts	11.045 Hochschulgruppen comments	11.045 Innovation, Entrepreneurs, Entwicklung posts
Institute, Fachbereiche comments	10.863 Arbeitskreise posts	11.045 Innovation, Entrepreneurs, Entwicklung posts	11.045 Uni Sports posts	11.045 Karriere Beruftseinstieg posts	11.136 Innovation, Entrepreneurs, Entwicklung comments
Karriere, Berufseinstieg comments	10.583 Musik comments	10.770 Rund um die Bibliothek	10.954 Musik posts	11.045 Fachschaften posts	11.045 Karriere Beruftseinsti

		comments			eg posts
KIT allegemein comments	10.863 Uni Sports posts	10.863 Arbeitskreise posts	10.954 Institute, Fachbereiche posts	11.045 Fachschaften comments	11.136 Hochschulgruppen comments
Musik comments	10.488 Musik posts	10.583 Karriere, Berufseinsteig comments	10.863 Rund um die Bibliothek comments	11.045 Karriere, Berufseinstieg posts	11.136 Fachschaften posts
Rund um die Bibliothek comments	10.770 Karriere, Berufseinsteig comments	10.863 Musik comments	10.954 Musik posts	11.045 Karriere, Berufseinstieg posts	11.136 Fachschaften comments
Social comments	10.954 Fachschaften comments	11.045 Hochschulgruppen comments	11.136 Institute, Fachbereiche comments	11.136 Arbeitskreise comments	11.136 Hochschulpolitik posts
Uni Sports comments	10.954 KIT allgemein posts	11.136 Hochschulgruppen comments	11.136 Social posts	11.136 Uni Sports posts	11.136 Hochschulpolitik posts
Arbeitskreise posts	10.488 Karriere, Berufseinstieg posts	10.770 Institute, Fachbereiche posts	10.863 KIT allgemein posts	10.863 Hochschulgruppen posts	10.863 Hochschulpolitik posts
Fachschaften posts	10.770 Karriere, Berufseinstieg posts	10.863 KIT allgemein posts	10.863 Institute, Fachbereiche posts	10.863 Uni Sports posts	10.863 Innovation, Entrepreneurs, Entwicklung comments
Hochschulgruppen posts	10.392 Institute, Fachbereiche posts	10.488 KIT allgemein posts	10.677 Karriere, Berufseinstieg posts	10.863 Arbeitskreise posts	10.954 Fachschaften posts
Hochschulpolitik posts	10.770 Uni Sports posts	10.863 Arbeitskreise posts	10.954 Fachschaften posts	11.045 Rund um die Bibliothek posts	11.045 KIT allgemein posts
Innovation,	10.770	10.954	10.954	11.045	11.045

Entrepreneurs, Entwicklung posts	Rund um die Bibliothek posts	Uni Sports posts	Hochschulgruppen posts	Fachschaften comments	KIT allgemein posts
Institute, Fachbereiche posts	10.392 Hochschulgruppen posts	10.677 KIT allgemein posts	10.770 Arbeitskreise posts	10.863 Fachschaften posts	10.954 Fachschaften posts
Karriere, Berufseinstieg posts	10.488 Arbeitskreise posts	10.677 Hochschulgruppen posts	10.770 KIT allgemein posts	10.770 Fachschaften posts	10.863 Musik posts
KIT allgemein posts	10.488 Hochschulgruppen posts	10.677 Institute, Fachbereiche posts	10.770 Karriere, Berufseinstieg posts	10.863 Fachschaften posts	10.863 Arbeitskreise posts
Musik posts	10.488 Musik comments	10.863 Karriere, Berufseinstieg posts	10.954 Institute, Fachbereiche posts	10.954 Rund um die Bibliothek comments	10.954 Karriere, Berufseinstieg comments
Rund um die Bibliothek posts	10.770 Innovation, Entrepreneurs, Entwicklung posts	11.045 Hochschulpolitik posts	11.045 Institute, Fachbereiche posts	11.136 Uni Sports posts	11.136 Musik posts
Social posts	10.770 Innovation, Entrepreneurs, Entwicklung comments	11.045 Hochschulpolitik comments	11.045 Fachschaften posts	11.136 Uni Sports comments	11.136 Rund um die Bibliothek comments
Uni Sports posts	10.770 Hochschulpolitik posts	10.863 Fachschaften posts	10.863 Karriere, Berufseinstieg posts	10.863 KIT allegemein comments	10.954 Innovation, Entrepreneurs, Entwicklung posts

References

Ahn, Sang-Hoon, Young Jun Choi, Young-Mi Kim, Sang-hoon Ahn Young, and Jun Choi. 2011. "Static Numbers to Dynamic Statistics: Designing a Policy-Friendly Social Policy Indicator Framework." *Social Indicators Research* 108 (3): 387–400. doi:10.1007/s11205-011-9875-9.

Allen, Stuart M, Martin Chorley, Gualtiero Colombo, Eva Jaho, Merkourios Karaliopoulos, Ioannis Stavrakakis, and Roger M Whitaker. 2014. "Exploiting User Interest Similarity and Social Links for Micro-Blog Forwarding in Mobile Opportunistic Networks." *Pervasive and Mobile Computing* 11 (April). Elsevier B.V.: 106–31. doi:10.1016/j.pmcj.2011.12.003.

Alpers, Georg W, Andrew J. Winzelberg, Catherine Classen, Heidi Roberts, Parvati Dev, Cheryl Koopman, and C. Barr Taylor. 2005. "Evaluation of Computerized Text Analysis in an Internet Breast Cancer Support Group." *Computers in Human Behavior* 21: 361–76. doi:10.1016/j.chb.2004.02.008.

Anand, Sudhir, and Amartya Sen. 1994. "Sustainable Human Development: Concepts and Priorities." *UNDP Human Development Report Office 1994 Occasional Papers. Available at SSRN:* http://ssrn.com/abstract=2294664.

Anderson, Laurel, Amy Ostrom, Canan Corus, Raymond P Fisk, Andrew S Gallan, Mario Giraldo, Martin Mende, et al. 2013. "Transformative Service Research: An Agenda for the Future." *Journal of Business Research* 66 (8). Elsevier Inc.: 1203–10. doi:10.1016/j.jbusres.2012.08.013.

Angner, Erik. 2005. *Is It Possible to Measure Happiness? The Measurement-Theoretic Argument against Subjective Measures of Wellbeing*. Birmingham, USA.

Antin, Judd, and Ef Churchill. 2011. "Badges in Social Media: A Social Psychological Perspective." *CHI 2011 Workshop Gamification Using Game Design Elements in NonGame Contexts 2011*, 1–4. http://uxscientist.com/public/docs/uxsci_2.pdf\nhttp://uxscientist.com/?sort=post_date&page=7.

Argamon, Shlomo, Moshe Koppel, James Pennebaker, and Jonathan Schler. 2009. "Automatically Profiling the Author of an Anonymous Text." *Communications of the ACM* 52 (2): 119–23. doi:10.1145/1461928.1461959.

Auer, Matthew R. 2011. "The Policy Sciences of Social Media." *The Policy Studies Journal* 39 (4): 709–36.

Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. 2010. "SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining." *LREC* 10: 2200–2204.

Back, Mitja D, Juliane M Stopfer, Simine Vazire, Sam Gaddis, Stefan C Schmukle, Boris Egloff, and Samuel D Gosling. 2010. "Facebook Profiles Reflect Actual Personality, Not Self-Idealization." *Psychological Science* 21 (3): 372–74. doi:10.1177/0956797609360756.

Balahur, Alexandra, and Jesús M Hermida. 2012. “Extending the EmotiNet Knowledge Base to Improve the Automatic Detection of Implicitly Expressed Emotions from Text.” In *LREC*, 1207–14.

Ballas, Dimitris. 2013. “What Makes a ‘happy City’?” *Cities* 32 (July): S39–50. doi:10.1016/j.cities.2013.04.009.

Baumeister, R F, and M R Leary. 1995. “The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation.” *Psychological Bulletin* 117 (3): 497–529. doi:10.1037/0033-2909.117.3.497.

Bazarova, N. N., J. G. Taft, Y. H. Choi, and D. Cosley. 2012. “Managing Impressions and Relationships on Facebook: Self- Presentational and Relational Concerns Revealed Through the Analysis of Language Style.” *Journal of Language and Social Psychology*. doi:10.1177/0261927X12456384.

Belsley, David A. 1991. “A Guide to Using the Collinearity Diagnostics.” *Computer Science in Economics and Management* 4: 33–50.

Bentham, Jeremy. 1789. *An Introduction to the Principles and Morals of Legislation*. Reprinted. Oxford: Blackwell. doi:10.4324/9780203209493_Introduction.

Berber-Sardinha, Tony. 2000. “Comparing Corpora with WordSmith Tools: How Large Must the Reference Corpus Be ?” *WCC ’00 Proceedings of the Workshop on Comparing Corpora*, 7–13. doi:10.3115/1117729.1117731.

Bergkvist, Lars, and John R Rossiter. 2007. “The Predictive of Multiple-Item Versus Validity Measures of the Same Constructs.” *Journal of Marketing Research* 44 (2): 175–84.

Berinsky, Adam J., Gregory Huber, and Gabriel S. Lenz. 2012. “Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk.” *Political Analysis* 20 (3): 351–68. doi:10.1093/pan/mpr057.

Bertsch, Valentin, Wolf Fitchner, Margeret Hall, Tobias Schumaker, and Christof Weinhardt. 2015. “Service Requirements for Consumer Engagement in the German Energy Retail Market.” In *Proceedings of 2015 Quality in Service (Quis14)*.

Bhutan, Kingdom of. 1991. *Seventh Five Year Plan: National Environmental Action PLan*. Vol. 1. Kingdom of Bhutan.

———. 2012. *Gross National Happiness Index Explained in Detail*. Thimphu, Bhutan.

Bishop, Jonathan. 2007. “Increasing Participation in Online Communities: A Framework for Human–computer Interaction.” *Computers in Human Behavior* 23 (4): 1881–93. doi:10.1016/j.chb.2005.11.004.

Blanchflower, David G., and Andrew J. Oswald. 2008. “Is Well-Being U-Shaped over the Life Cycle?” *Social Science and Medicine* 66 (8): 1733–49.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3: 993–1022.

Böcking, Benedikt, Margeret Hall, and Jeff Schneider. 2015. “Event Prediction With Learning Algorithms—A Study of Events Surrounding the Egyptian Revolution of 2011 on the Basis of Micro Blog Data.” *Policy & Internet*. doi:10.1002/poi3.89.

Bosnjak, Michael, and Tracy L Tuten. 2001. “Classifying Response Behaviors in Web-Based Surveys.” *Journal of Computer-Mediated Communication* 6 (3): 14.

Boulianne, Shelley. 2009. “Does Internet Use Affect Engagement? A Meta-Analysis of Research.” *Political Communication* 26 (January 2015): 193–211. doi:10.1080/10584600902854363.

Braga-Neto, Ulisses M, and Edward R Dougherty. 2004. “Is Cross-Validation Valid for Small-Sample Microarray Classification?” *Bioinformatics* 20 (3): 374–80. doi:10.1093/bioinformatics/btg419.

Brown, Penelope, and Stephen Levinson. 2013. *Politeness Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Buckels, Erin E., Paul D. Trapnell, and Delroy L. Paulhus. 2014. “Trolls Just Want to Have Fun.” *Personality and Individual Differences* 67 (September). Elsevier Ltd: 97–102. doi:10.1016/j.paid.2014.01.016.

Burke, Moira, Cameron Marlow, and Thomas Lento. 2009. “Feed Me : Motivating Newcomer Contribution in Social Network Sites.” In *CHI 2009*, 1–10. Boston, US.

———. 2010. “Social Network Activity and Social Well-Being.” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1909–12. doi:10.1145/1753326.1753613.

Burnap, Pete, Omer Rana, Matthew Williams, William Housley, Adam Edwards, Jeffrey Morgan, Luke Sloan, and Javier Conejero. 2014. “COSMOS: Towards an Integrated and Scalable Service for Analysing Social Media on Demand.” *International Journal of Parallel, Emergent and Distributed Systems*, no. January 2015: 37–41. doi:10.1080/17445760.2014.902057.

Burnap, Pete, and Matthew Williams. 2014. “Hate Speech , Machine Classification and Statistical Modelling of Information Flows on Twitter : Interpretation and Communication for Policy Decision Making.” In , 1–18.

Burrows, R., and M. Savage. 2014. “After the Crisis? Big Data and the Methodological Challenges of Empirical Sociology.” *Big Data & Society* 1 (1). doi:10.1177/2053951714540280.

Calvo, Rafael A, and Sidney D Mello. 2010. “Affect Detection : An Interdisciplinary Review of Models , Methods , and Their Applications.” *IEEE Transactions on Affective Computing* 1 (1): 18–37.

Cameron, Kim S., David Bright, and Arran Caza. 2004. “Exploring the Relationships between Organizational Virtuousness and Performance.” *American Behavioral Scientist* 47 (6): 766–90. doi:10.1177/0002764203260209.

Caspi, Avner, and Paul Gorsky. 2006. “Online Deception: Prevalence, Motivation, and Emotion.” *Cyberpsychology & Behavior* 9 (1): 54–60.

Catanese, Salvatore, Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, and Alessandro Provetti. 2011. “Crawling Facebook for Social Network Analysis Purposes.” *Proceedings of the International Conference on Web Intelligence, Mining and Semantics - WIMS ’11*. New York, New York, USA: ACM Press, 1. doi:10.1145/1988688.1988749.

Caton, Simon, Lukas Brückner, Margeret Hall, and Christof Weinhardt. 2015. “FBWatch: Extracting , Analyzing and Visualizing Public Facebook Profiles.” Karlsruhe, Germany.

Caton, Simon, Margeret Hall, and Christof Weinhardt. 2015. “How Do Politicians Use Facebook? An Applied Social Observatory.” *Big Data & Society*.

Chou, Y.-T., and W.-C. Wang. 2010. “Checking Dimensionality in Item Response Models With Principal Component Analysis on Standardized Residuals.” *Educational and Psychological Measurement* 70 (5): 717–31. doi:10.1177/0013164410379322.

Chung, Cindy, and James Pennebaker. 2007. “The Psychological Functions of Function Words.” Edited by K Fiedler. *Social Communication*. New York, New York, USA: Psychology Press, 343–59. doi:10.4324/9780203837702.

———. 2011. “Linguistic Inquiry and Word Count (LIWC): Pronounced like ‘Luke’ ... and Other Useful Facts.” In *Applied Natural Language Processing*, edited by Cyrus Shaoul and Chris Westbury, 206–8. doi:10.4018/978-1-60960-741-8.

———. 2014. “Counting Little Words in Big Data: The Psychology of Communities, Culture, and History.” In *Social Cognition and Communication*, edited by Joseph Forgas, Orsolya Vincze, and Janos Laszlo, 25–42. New York, New York, USA: Psychology Press.

Chung, Jessica, and Eni Mustafaraj. 2010. “Can Collective Sentiment Expressed on Twitter Predict Political Elections ?” In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 1770–71.

Cioffi-Revilla, Claudio. 2010. “Computational Social Science.” *Computational Statistics* 2 (3): 259–71. doi:10.1002/wics.95.

———. 2014. *Introduction to Computational Social Science*. Berlin: Springer Texts in Computer Science.

Clark, Andrew E, Paul Frijters, and Michael A Shields. 2008. “Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles.” *Journal of Economic Literature* 46 (1): 95–144.

Commission, European. 2011. *Well-Being Aggregate Report September 2011 Eurobarometer Qualitative Studies*.

Conway, James M., and Charles E Lance. 2010. “What Reviewers Should Expect from Authors Regarding Common Method Bias in Organizational Research.” *Journal of Business and Psychology* 25 (3): 325–34. doi:10.1007/s10869-010-9181-6.

Coviello, By Lorenzo, James H Fowler, and Massimo Franceschetti. 2014. “Words on the Web: Noninvasive Detection of Emotional Contagion in Online Social Networks.” *Proceedings of the IEEE* 102 (12): 1911–21.

Danescu-Niculescu-Mizil, Cristian, Michael Gamon, and Susan Dumais. 2011. “Mark My Words! Linguistic Style Accommodation in Social Media.” In *WWW 2011*. Hyperabad, India: ACM Press. doi:10.1145/1963405.1963509.

Das, Sauvik, and Adam Kramer. 2013. “Self-Censorship on Facebook.” *Association for the Advancement of Artificial Intellegence*, 1–8.

Davenport, Shaun W., Shawn M. Bergman, Jacqueline Z. Bergman, and Matthew E. Fearrington. 2014. “Twitter versus Facebook: Exploring the Role of Narcissism in the Motives and Usage of Different Social Media Platforms.” *Computers in Human Behavior* 32 (March). Elsevier Ltd: 212–20. doi:10.1016/j.chb.2013.12.011.

Davies, James C. 1962. “Towards a Theory of Revolution.” *American Sociological Review* 27 (1): 5–19.

Deci, Edward L., and Richard M. Ryan. 2006. “Hedonia, Eudaimonia, and Well-Being: An Introduction.” *Journal of Happiness Studies* 9 (1): 1–11. doi:10.1007/s10902-006-9018-1.

———. 2008. “Self-Determination Theory: A Macrotheory of Human Motivation, Development, and Health.” *Canadian Psychology/Psychologie Canadienne* 49 (3): 182–85. doi:10.1037/a0012801.

DeNeve, Kristina M., and Harris Cooper. 1998. “The Happy Personality: A Meta-Analysis of 137 Personality Traits and Subjective Well-Being.” *Psychological Bulletin* 124 (2): 197–229.

Deterding, Sebastian. 2011. “Situated Motivational Affordances of Game Elements : A Conceptual Model.” *CHI 2011 Workshop Gamification Using Game Design Elements in NonGame Contexts 2011*, 3–6. doi:ACM 978-1-4503-0268-5/11/05.

Deterding, Sebastian, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. “From Game Design Elements to Gamefulness.” In *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek ’11*, 9–11. doi:10.1145/2181037.2181040.

Diener, Ed. 1984a. “Subjective Well-Being.” *Psychological Bulletin* 95 (3): 542–75.

———. 1984b. “Subjective Well-Being: The Science of Happiness and a Proposal for a National Index.” *American Psychologist* 55 (1): 34–43. doi:10.1037//0003-066X.55.1.34.

———. 1994. “Assessing Subjective Well-Being: Progress and Oppurtunities.” *Social Indicators Research* 31: 103–57.

———. 2006. “Guidelines for National Indicators of Subjective Well-Being and Ill-Being.” *Applied Research in Quality of Life* 1 (2): 151–57. doi:10.1007/s11482-006-9007-x.

Diener, Ed, and Micaela Chan. 2011. “Happy People Live Longer: Subjective Well-Being Contributes to Health and Longevity.” *Applied Psychology: Health and Well-Being* 3 (1): 1–43.

Diener, Ed, Robert A Emmons, Randy Larson, and Sharon Griffin. 1985. “Satisfaction with Life Scale.” *Journal of Personality Assessment* 49 (1): 71–75.

Diener, Ed, and Martin Seligman. 2002. “Very Happy People.” *Psychological Science* 13 (1): 81–84.

———. 2004. “Toward an Economy of Well-Being.” *Psychological Science in the Public Interest* 5 (1): 1–31.

Diener, Ed, and Eunkook Suh. 1997. “Measuring Quality of Life: Economic, Social, and Subjective Indicators.” *Social Indicators Research* 40: 189–216.

Diener, Ed, Eunkook Suh, Richard E Lucas, and Heidi Smith. 1999. “Subjective Well-Being: Three Decades of Progress.” *Psychological Bulletin* 125 (2): 276–302.

Dimitrova, Daniela, Adam Shehata, Jesper Stromback, and Lars Nord. 2011. “The Effects of Digital Media on Political Knowledge and Participation in Election Campaigns: Evidence From Panel Data.” *Communication Research* 41 (1): 95–118. doi:10.1177/0093650211426004.

Dixon, Dan. 2011. “Player Types and Gamification.” *CHI 2011 Workshop Gamification Using Game Design Elements in NonGame Contexts*, 12–15. doi:ACM 978-4503-0268-5/11/05.

Dodds, Peter Sheridan, Kameroncker Decker Harris, Isabel M. Kloumann, Catherine a. Bliss, and Christopher M. Danforth. 2011. “Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter.” *PLoS ONE* 6 (12). doi:10.1371/journal.pone.0026752.

Dodge, Rachel, Annette P Daly, Jan Huyton, and Lalage D Sanders. 2012. “The Challenge of Defining Wellbeing.” *International Journal of Well-Being* 2 (3): 222–35. doi:10.5502/ijw.v2i3.4.

Drezner, Daniel W. 2004. “The Global Governance of the Internet: Bringing the State Back In.” *Political Science Quarterly* 119 (3): 477–98.

Dworman, G., Steven O. Kimbrough, and J.D. Laing. 1995. “On Automated Discovery of Models Using Genetic Programming in Game-Theoretic Contexts.” In *Proceedings of the Twenty-Eighth Annual Hawaii International Conference on System Sciences*, 3:428–38. doi:10.1109/HICSS.1995.375625.

Easterlin, Richard A. 1974. “Does Economic Growth Improve the Human Lot? Some Empirical Evidence.” In *Nations and Households in Economic Growth*, edited by Paul A. David and Melvin W. Reder, 89–125. New York: Academic Press, Inc.

———. 1995. “Will Raising the Incomes of All Increase the Happiness of All ?” *Journal of Economic Behavior and Organization* 27: 35–47.

Easterlin, Richard A, Laura Angelescu McVey, Malgorzata Switek, Onnicha Sawangfa, and Jacqueline Smith Zweig. 2010. “The Happiness-Income Paradox Revisited.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (52): 22463–68. doi:10.1073/pnas.1015962107.

Ellison, Nicole, Rebecca Heino, and Jennifer Gibbs. 2006. “Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment.” *Journal of Computer-Mediated Communication* 11 (2): 415–41. doi:10.1111/j.1083-6101.2006.00020.x.

Escher, Tobias. 2013. “Mobilisierung Zu Politischer Partizipation Durch Das Internet: Erwartungen, Erkenntnisse Und Herausforderungen Der Forschung.” *Analyse & Kritik* 35: 449–76.

Ewig, Caterina. 2011. “Social Media: Theorie Und Praxis Digitaler Sozialität / Social Media : Theory and Practice of Digital Sociality.” In *Social Media: Theorie Und Praxis Digitaler Sozialität*, edited by Mario Anastasiadis and Caja Thimm. Frankfurt am Main: Peter Lang Internationaler Verlag der Wissenschaten.

Fiesler, Casey, and Amy Bruckman. 2014. “Copyright Terms in Online Creative Communities.” In *CHI ’14 Extended Abstracts on Human Factors in Computing Systems*, 2551–56. Toronto, CA: ACM Press. doi:10.1145/2559206.2581294.

Fowler, James, and Nicholas Christakis. 2008. “Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis over 20 Years in the Framingham Heart Study.” *BMJ (Clinical Research Ed.)* 337: a2338. doi:10.1136/bmj.a2338.

Fowler, Robert L. 1987. “Power and Robustness in Product-Moment Correlation.” *Applied Psychological Measurement* 11 (4): 419–28.

Frey, Bruno S, and Jana Gallus. 2013. “Subjective Well-Being and Policy.” *Topoi* DOI 10.100 (April): 1–6. doi:10.1007/s11245-013-9155-1.

Frey, Bruno S, and Alois Stutzer. 2001. *Happiness and Economics: How the Economy and Institutions Affect Human Well-Being*. Princeton, NJ: Princeton UP.

———. 2007. *Should National Happiness Be Maximized?* 306. *Research In Economics*. Vol. 306. 1424-0459. Zurich.

———. 2012. *Recent Developments in the Economics of Happiness: A Selected Overview*. 7078. Discussion Paper Series. Bonn, Germany. http://hdl.handle.net/10419/69369.

Friedman, Batya. 1996. “Value-Sensitive Design.” *Interactions* 3: 16–23. doi:10.1145/242485.242493.

Friedman, Batya, Peter H Kahn Jr., and Alan Borning. 2003. “Value Sensitive Design and Information Systems.” In *The Handbook of Information and Computer Ethics*, edited by Kenneth Einar Kima and Herman T Tavani, 69–102. Hoboken, NJ: Wiley. doi:10.4018/978-1-59140-144-5.

Galesic, M., and Michael Bosnjak. 2009. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." *Public Opinion Quarterly* 73 (2): 349–60. doi:10.1093/poq/nfp031.

Gasper, Des. 2005. "Subjective and Objective Well-Being in Relation to Economic Inputs: Puzzles and Responses." *Review of Social Economy* 63 (2): 177–206. doi:10.1080/00346760500130309.

Gebauer, Heiko, and Javier Reynoso. 2013. "An Agenda for Service Research at the Base of the Pyramid." *Journal of Service Management* 24 (5): 482–502.

Giles, Howard, Nikolas Coupland, and Justine Coupland. 1991. "Accomodation Theory: Communication, Context, and Consequence." In *Contexts of Accomodation: Developments in Applied Sociolinguistics*, edited by Howard Giles, Nikolas Coupland, and Justine Coupland, 1–68. Cambridge University Press.

Gleser, Leon J. 1966. "A Noted on the Sphericity Test." *The Annuals of Mathematical Statistics* 37 (2): 464–67.

Go, Alec, Richa Bhayani, and Lei Huang. 2009. "Twitter Sentiment Classification Using Distant Supervision." Stanford, USA.

Goffman, Erving. 1959. *The Presentation of Self In Everyday Life*. 1st ed. New York, New York, USA: Anchor.

Golder, Scott A, and Michael W Macy. 2012. "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures." *Science* 333: 1878–81. doi:10.1126/science.1202775.

Gonzales, Amy L., Jeffrey T. Hancock, and James Pennebaker. 2010. "Language Style Matching as a Predictor of Social Dynamics in Small Groups." *Communication Research* 37 (1): 3–19. doi:10.1177/0093650209351468.

González-Bailón, Sandra, Ning Wang, Alejandro Rivero, and Javier Borge-Holthoefer. 2014. "Assessing the Bias in Samples of Large Online Networks." *Social Networks* 38 (January). Elsevier B.V.: 16–27. doi:10.1016/j.socnet.2014.01.004.

Grant, Adam M., Marlys K. Christianson, and Richard H. Price. 2007. "Happiness, Health, or Relationships? Managerial Practices and Employee Well-Being Tradeoffs." *Academy of Management Perspectives* 21 (3): 51–63. doi:10.5465/AMP.2007.26421238.

Grawitch, Matthew J., Melanie Gottschalk, and David C. Munz. 2006. "The Path to a Healthy Workplace: A Critical Review Linking Healthy Workplace Practices, Employee Well-Being, and Organizational Improvements." *Consulting Psychology Journal: Practice and Research* 58 (3): 129–47. doi:10.1037/1065-9293.58.3.129.

Groom, Carla J, and James Pennebaker. 2002. "Words." *Journal of Research in Personality* 36: 615–21. doi:10.1016/S0092-6566(02)00512-3.

Gunsch, Mark A, Sheila Brownlow, Sarah E Haynes, and Zachary Mabe. 2000. "Differential Forms Linguistic Content of Various of Political Advertising." *Journal of Broadcasting and Electronic Media* 44 (1): 27–42. doi:10.1207/s15506878jobem4401.

Guven, Cahit, and Bent E Sørensen. 2012. "Subjective Well-Being : Keeping Up with the Perception of the Joneses." *Social Indicators Research* 109: 439–69. doi:10.1007/s11205-011-9910-x.

Haas, Christian, Simon Caton, and Christof Weinhardt. 2011. "Engineering Incentives in Social Clouds." In *Proceedings - 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGrid 2011*, 572–75. doi:10.1109/CCGrid.2011.52.

Hackenberg, Robert. 1970. "The Social Observatory : Time Series Data for Health and Behavioral Research." *Social Science and Medicine* 4: 343–57.

Hall, Margeret, and Simon Caton. 2014. *A Crowdsourcing Approach to Identify Common Method Bias and Self-Representation*. *IPP2014: Crowdsourcing for Politics and Policy*. The Internet, Policy, and Politics. Oxford, England.

Hall, Margeret, Simon Caton, and Christof Weinhardt. 2013. "Well-Being's Predictive Value." In *Proceedings of the 15th International Conference on Human-Computer Interaction (HCII)*, edited by A. A. Ozok and P. Zaphiris, 13–22. Berlin: LNCS, Springer Verlag.

Hall, Margeret, Stefan Glanz, Simon Caton, and Christof Weinhardt. 2013. "Measuring Your Best You: A Gamification Framework for Well-Being Measurement." In *Third International Conference on Social Computing and Its Applications*, 277–82. Karlsruhe, Germany: IEEE. doi:doi:10.1109/CGC.2013.51.

Hall, Margeret, Christian Haas, Steven O. Kimbrough, and Christof Weinhardt. 2014. "An Extended Conceptual Framework for Transformative Service Research." In *AMA SERVSIG International Service Research Conference*, edited by Rodoula Tsiotsou, forthcoming. Thessaloniki, GR.

Hall, Margeret, Steven O. Kimbrough, Christian Haas, Christof Weinhardt, and Simon Caton. 2012. "Towards the Gamification of Well-Being Measures." In *2012 IEEE 8th International Conference on E-Science, E-Science 2012*, 1–8. Ieee. doi:10.1109/eScience.2012.6404457.

Hampton, Keith N, Lee Rainie, Weixu Lu, Maria Dwyer, Inyoung Shin, and Kristen Purcell. 2014. *Social Media and the Spiral of Silence*. Washington D.C.

Hampton, Keith N, Lauren Sessions Goulet, Lee Rainie, and Kristen Purcell. 2011. *Social Networking Sites and Our Lives*. Washington D.C.

Hancock, Jeffrey. 2007. "Digital Deception: Why, When, and How People Lie Online." In *Oxford Handbook of Internet Psychology*, edited by Adam Joinson, Katelyn McKenna, Tom Postmes, and Ulf-Dietrich Reips. Oxford, England.

Hancock, Jeffrey T., Christopher Landrigan, and Courtney Silver. 2007. "Expressing Emotion in Text-Based Communication." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07*, 929–32. doi:10.1145/1240624.1240764.

Harris, Sam. 2010. *The Moral Landscape: How Science Can Determine Human Values*. Free Press.

Harter, James K, Frank L Schmidt, and Corey L M Keyes. 2003. "Well-Being in the Workplace and Its Relationship to Business Outcomes: A Review of the Gallup Studies." In *Flourishing: The Positive Person and the Good Life*, edited by Corey L. M. Keyes and Jonathan Haidt, 205–25. Washington D.C.: American Psychological Association.

Haslam, Nick, Jennifer Whelan, and Brock Bastian. 2009. "Big Five Traits Mediate Associations between Values and Subjective Well-Being." *Personality and Individual Differences* 46 (1). Elsevier Ltd: 40–42.

Hayfield, Tristen, and Jeffrey S. Racine. 2013. *Package "np."* 0.50-1. R-project.org.

Hazelkorn, Ellen, Tia Loukkola, and Thérèse Zhang. 2014. *Rankings in Institutional Strategies and Processes: Impact or Illusion?* Brussels.

Hevner, Alan R. 2007. "A Three Cycle View of Design Science Research." *Scandinavian Journal of Information Systems* 19 (2): 87–92.

Hevner, Alan R, Salvatore T March, Jinsoo Park, and Sudha Ram. 2004. "Design Science in Information Systems Research." *MIS Quarterly* 28 (1): 75–105.

Hilsen, Anne Inga, and Tove Helvik. 2012. "The Construction of Self in Social Medias, such as Facebook." *AI & Society* 29 (1): 3–10. doi:10.1007/s00146-012-0426-y.

Hirschauer, Norbert, Mira Lehberger, and Oliver Musshoff. 2014. "Happiness and Utility in Economic Thought—Or: What Can We Learn from Happiness Research for Public Policy Analysis and Public Policy Making?" *Social Indicators Research*, June. doi:10.1007/s11205-014-0654-2.

Hoever, André. 2010. "Strategien Und Konzepte Der Selbstdarstellung Auf Social Network Services Am Beispiel Facebook." Berlin: Berliner Methodentreffen Qualitative Forschung.

Hogan, B. 2010. "The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online." *Bulletin of Science, Technology & Society* 30 (6): 377–86. doi:10.1177/0270467610385893.

Holmes, Danielle, Georg W Alpers, Tasneem Ismailji, Catherine Classen, Talor Wales, Valerie Cheasty, Andrew Miller, and Cheryl Koopman. 2007. "Cognitive and Emotional Processing in Narratives of Women Abused by Intimate Partners." *Violence Against Women* 13 (11): 1192–1205. doi:10.1177/1077801207307801.

Housley, William, Rob Procter, Adam Edwards, Pete Burnap, Matthew Williams, Luke Sloan, Omer Rana, Jeffrey Morgan, Alex Voss, and Anita Greenhill. 2014. "Big and Broad Social Data and the Sociological Imagination: A Collaborative Response." *Big Data & Society* 1 (2). doi:10.1177/2053951714545135.

Hsee, Christopher K., Reid Hastie, and Jingqiu Chen. 2008. “Hedonomics: Bridging Decision Research With Happiness Research.” *Perspectives on Psychological Science* 3 (3): 224–43. doi:10.1111/j.1745-6924.2008.00076.x.

Hsieh, Hsiu-Fang, and Sarah E Shannon. 2005. “Three Approaches to Qualitative Content Analysis.” *Qualitative Health Research* 15 (9): 1277–88. doi:10.1177/1049732305276687.

Huotari, Kai, and Juho Hamari. 2012. “Defining Gamification - A Service Marketing Perspective.” In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 17–22. doi:10.1145/2393132.2393137.

Huppert, Felicia, and Timothy T C So. 2009. *What Percentage of People in Europe Are Flourishing and What Characterises Them?* Florence, Italy. http://www.isqols2009.istitutodeglinnocenti.it/Content_en/Huppert.pdf.

———. 2013. “Flourishing Across Europe: Application of a New Conceptual Framework for Defining Well-Being.” *Social Indicators Research* 110 (3): 837–61. doi:10.1007/s11205-011-9966-7.

IBM. 2011a. *IBM SPSS Regression 22*.

———. 2011b. *IBM SPSS Advanced Statistics 22*.

Jaho, Eva, Merkourios Karaliopoulos, and Ioannis Stavrakakis. 2011. “ISCoDe: A Framework for Interest Similarity-Based Community Detection in Social Networks.” *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April. Ieee, 912–17. doi:10.1109/INFCOMW.2011.5928942.

Johann, Holger, Margeret Hall, Steven O. Kimbrough, Nicholas Quintus, and Christof Weinhardt. 2014. “Service District Optimization.” *SSRN Electronic Journal*. doi:http://dx.doi.org/10.2139/ssrn.2438247.

John, Oliver P., Eileen M. Donahue, and Robert L. Kentle. 1991. *The Big Five Inventory—versions 4a and 54*. Berkeley, USA.

Joiner Jr., Thomas. 2005. *Why People Die By Suicide: Further Development and Tests of the Interpersonal-Psychological Theory of Suicidal Behavior*. Cambridge, USA: Harvard University Press.

Joiner, Thomas E., Daniel Hollar, and Kimberly Van Orden. 2006. “On Buckeyes, Gators, Super Bowl Sunday, and the Miracle on Ice: ‘Pulling Together’ Is Associated With Lower Suicide Rates.” *Journal of Social and Clinical Psychology* 25 (2): 179–95. doi:10.1521/jscp.2006.25.2.179.

Jowell, Roger, Caroline Roberts, Rory Fitzgerald, and Gillian Eva. 2006. *Meausring Attitudes Cross-Nationally: Lessons Form the European Social Survey*. Vol. 2006. London: SAGE Publications.

Jungherr, Andreas, Pascal Jürgens, and Harald Schön. 2011. “Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. 'Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment.'" *Social Science Computer Review* 30 (2): 229–34. doi:10.1177/0894439311404119.

Kahneman, Daniel. 2009. "New Challenges to the Rationality Assumption." *Legal Thoery* 3 (2): 1997. doi:10.1017/S1352325200000689.

Kahneman, Daniel, and Alan B Krueger. 2006. "Developments in the Measurement of Subjective Well-Being." *Journal of Economic Perspectives* 20 (1): 3–24.

Kahneman, Daniel, Alan B Krueger, David Schkade, Norbert Schwarz, and Arthur Stone. 2004a. "Toward National Well-Being Accounts." *The American Economic Review* 94 (2): 429–34.

———. 2004b. "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method." *Science* 306 (5702): 1776–80. doi:10.1126/science.1103572.

———. 2006. "Would You Be Happier If You Were Richer? A Focusing Illusion." *Science* 312 (5782): 1908–10. doi:10.1126/science.1129688.

Kahneman, Daniel, and Richard H Thaler. 2006. "Anomalies Utility Maximization and Experienced Utility." *Journal of Economic Perspectives* 20 (1): 221–34.

Kaiser, Henry F. 1970. "A Second Generation Little Jiffy." *Psychometrika* 35 (4): 401–15.

Kaplan, Andreas M., and Michael Haenlein. 2010. "Users of the World, Unite! The Challenges and Opportunities of Social Media." *Business Horizons* 53: 59–68. doi:10.1016/j.bushor.2009.09.003.

Karlsson, Niklas, George Loewenstein, and Jane McCafferty. 2004. "Economics of Meaning." *Nordic Journal of Political Economy* 30 (1): 61–75.

Kassarjian, Harold H. 1977. "Content Analysis in Consumer Research." *Journal of Consumer Research* 4 (1): 8–18.

Killingsworth, Matthew, and Daniel T Gilbert. 2010. "A Wandering Mind Is an Unhappy Mind." *Science (New York, N.Y.)* 330: 932. doi:10.1126/science.1192439.

Kim, Soo-min, Patrick Pantel, Tim Chklovski, and Marco Pennacchiotti. 2006. "Automatically Assessing Review Helpfulness." In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP*, 423–30. Sydney, Australia.

Kivelä, Aki, and Olli Lyytinen. 2004. "Topic Map Aided Publishing – A Case Study of Assembly Media Archive." In *STeP 2004 - The 11th Finnish Artificial Intelligence Conference Proceedings*.

Komisin, Mike, and Curry Guinn. 2012. "Identifying Personality Types Using Document Classification Methods." In *Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference*, 232–37.

Kouloumpis, Efthymios, Theresa Wilson, and Johanna Moore. 2011. “Twitter Sentiment Analysis : The Good the Bad and the OMG !,” 538–41.

Kramer, Adam. 2010. “An Unobtrusive Behavioral Model of ‘Gross National Happiness.’” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 287–90. doi:10.1145/1753326.1753369.

———. 2012. “The Spread of Emotion via Facebook.” In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI ’12*, 767–70. New York, New York, USA: ACM Press. doi:10.1145/2207676.2207787.

Kramer, Adam, Susan R. Fussel, Leslie D. Setlock, Susan R Fussell, and Leslie D. Setlock. 2004. “Text Analysis as a Tool for Analyzing Conversation in Online Support Groups.” *Proceedings of the Conference on Human Factors in Computing Systems 2004*. Vienna, Austria, 1485–88. doi:10.1145/985921.986096.

Kramer, Adam, J. E. Guillory, and Jeffrey Hancock. 2014. “Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks.” *Proceedings of the National Academy of Sciences*, June. doi:10.1073/pnas.1320040111.

Kross, Ethan, Philippe Verduyn, Emre Demiralp, Jiyoung Park, David Seungjae Lee, Natalie Lin, Holly Shablack, John Jonides, and Oscar Ybarra. 2013. “Facebook Use Predicts Declines in Subjective Well-Being in Young Adults.” *PLOS One* 8 (8): e69841. doi:10.1371/journal.pone.0069841.

Kumar, AK Shiva, Zenda Munro Ofir, Kunzang Dechen Dorji, Ruth Abraham, and Elizabeth K. Lang. 2007. *Assessment of Development Results: Bhutan*. New York, New York, USA.

Kushin, Matthew James, and Masahiro Yamamoto. 2010. “Did Social Media Really Matter? College Students’ Use of Online Media and Political Decision Making in the 2008 Election.” *Mass Communication and Society* 13 (5): 608–30. doi:10.1080/15205436.2010.516863.

Lacey, Heather P, Angela Fagerlin, George Loewenstein, Dylan M Smith, Jason Riis, and Peter Ubel. 2008. “Are They Really That Happy? Exploring Scale Recalibration in Estimates of Well-Being.” *Health Psychology* 27 (6): 669–75. doi:10.1037/0278-6133.27.6.669.

Lance, Charles E, Marcus M Butts, and Lawrence C Michels. 2006. “The Sources of Four Commonly Reported Cutoff Criteria: What Did They Really Say ?” *Organizational Research Methods* 9 (2): 202–20. doi:10.1177/1094428105284919.

Larsson, A, T Larsson, L Leifer, J Feland, M Van der Loos, and J Feland. 2005. “Design for Wellbeing: Innovations for People.” *International Conference on Engineering Design (ICED05)*, 1–10.

Lasswell, Harold D. 1967. “Do We Need Social Observatories?” *The Saturday Review*.

Lawson, Helene M., and Kira Leck. 2006. “Dynamics of Internet Dating.” *Social Science Computer Review* 24 (2): 189–208. doi:10.1177/0894439305283402.

Lazer, David, Devon Brewer, Nicholas Christakis, James Fowler, and Gary King. 2009. “Life in the Network: The Coming Age of Computational Social.” *Science* 323 (5915): 721–23. doi:10.1126/science.1167742.Life.

Lin, Han, and Lin Qiu. 2013. “Two Sites, Two Voices: Linguistic Differences between Facebook Status Updates and Tweets.” Edited by P.L.P. Rau. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8024 LNCS. LNCS 8024: 432–40. doi:10.1007/978-3-642-39137-8-48.

Lindner, Andreas, Claudia Niemeyer, Simon Caton, Margeret Hall, Claudia Niemeyer, and Simon Caton. 2015. “BeWell: A Sentiment Aggregator for Proactive Community Management.” In *CHI’15 Extended Abstracts*. doi:http://dx.doi.org/10.1145/2702613.2732787.

Lingel, Jessa, Mor Naaman, and danah boyd. 2014. “City, Self, Network: Transnational Migrants and Online Identity Work.” In *CSCW’14*, 1502–10. doi:10.1145/2531602.25311693.

Linville, Patricia W. 1985. “Self-Complexity and Affective Extremity: Don’t Put All of Your Eggs in One Cognitive Basket.” *Social Cognition* 3 (1): 94–120.

Liu, Bing. 2010. “Sentiment Analysis and Subjectivity.” In *Handbook of Natural Language Processing*, edited by N. Indurkhya and F.J. Damerau, 1–38.

Lyubomirsky, Sonja, Laura King, and Ed Diener. 2005. “The Benefits of Frequent Positive Affect: Does Happiness Lead to Success?” *Psychological Bulletin* 131 (6): 803–55. doi:10.1037/0033-2909.131.6.803.

Mahmud, Jalal. 2014. “Why Do You Write This ? Prediction of Influencers from Word Use.” In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 603–6.

Margetts, Helen, Peter John, Tobias Escher, and Stéphane Reissfelder. 2011. “Social Information and Political Participation on the Internet: An Experiment.” *European Political Science Review* 3 (3): 321–44. doi:10.1017/S1755773911000129.

Markham, Annette, and Elizabeth Buchanan. 2012. “Ethical Decision-Making and Internet Research.” Association of Internet Researchers.

McCallum, Andrew Kachites. 2002. “MALLET: A Machine Learning for Language Toolkit.” http://mallet.cs.umass.edu.

McCullagh, Peter. 1984. “Generalized Linear Models.” *European Journal of Operations Research* 16: 285–92.

Mckelvey, Karissa. 2013. “Truthy : Enabling the Study of Online Social Networks.” In *CSCW’13*, 23–25. San Antonio, TX: ACM Press.

Mehl, Matthias R, and James Pennebaker. 2003. “The Sounds of Social Life: A Psychometric Analysis of Students’ Daily Social Environments and Natural Conversations.” *Journal of Personality and Social Psychology* 84 (4): 857–70. doi:10.1037/0022-3514.84.4.857.

Mehra, Ajay, Martin Kilduff, and Daniel J Brass. 2001. “The Social Networks of High and Low Self-Monitors: Implications for Workplace Performance.” *Administrative Science Quarterly* 46 (1): 121–46.

Meltzer, Marisa. 2014. “Hashtags in Titles Is a # Trend That Can Backfire.” *New York Times*, October 8. http://www.nytimes.com/2014/10/09/fashion/hashtags-in-titles-is-a-trend-that-can-backfire.html.

Mende, Martin, Ruth N Bolton, and Mary Jo Bitner. 2013. “Decoding Customer-Firm Relationships: How Attachment Styles Help Explain Customers’ Preferences for Closeness, Repurchase Intentions, and Changes in Relationship Breadth.” *Journal of Marketing Research* L (February): 125–42.

Mitchell, Lewis, Morgan R Frank, Kameron Decker Harris, Peter Sheridan Dodds, and Christopher M Danforth. 2013. “The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place.” *PLOS One* 8 (5): e64417. doi:10.1371/journal.pone.0064417.

NEF. 2009. *National Accounts of Well-Being*. London.

Nelder, J. A., and R. W. M. Wedderburn. 1972. “Generalized Linear Models.” *Journal of the Royal Statistical Society. Series A (General)* 135 (3): 370–84.

Newman, Matthew, Carla J. Groom, Lori D. Handelman, and James Pennebaker. 2008. “Gender Differences in Language Use: An Analysis of 14,000 Text Samples.” *Discourse Processes* 45: 211–36. doi:10.1080/01638530802073712.

Newman, Matthew, James Pennebaker, Diane Berry, and Jane Richards. 2003. “Lying Words: Predicting Deception From Linguistic Styles.” *Personality and Social Psychology Bulletin* 29: 665–75. doi:10.1177/0146167203251529.

Niederhoffer, Kate, and James Pennebaker. 2002. “Linguistic Synchrony in Social Interaction.” *Journal of Language and Social Psychology* 21 (4). Austin: 337–60.

Noelle-Neumann, Elisabeth. 1974. “The Spiral of Silence: A Theory of Public Opinion.” *Journal of Communications* 24 (2): 43–51.

Norman, Wayne, and Chris MacDonald. 2004. “Getting to the Bottom of ‘Triple Bottom Line.’” *Business Ethics Quarterly* 14 (2): 243–62.

O’Connor, Brendan, Ramnath Balasubramanyan, Bryan R Routledge, Noah A Smith, Brendan O’Connor, Ramnath Balasubramanyan, Bryan R Routledge, and Noah A Smith. 2010. “From Tweets to Polls : Linking Text Sentiment to Public Opinion Time Series.” In *From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series*, 122–29.

Oberlander, Jon, and Scott Nowson. 2006. “Whose Thumb Is It Anyway? Classifying Author Personality from Weblob Text.” In *21st International Conference on Computational Linguistics*, 627–34. Sydney, Australia: Association for Computational Linguistics.

OECD. 2007. *Participative Web and User-Created Content*. *Participative Web and User-Created Content*. Brussels. doi:10.1787/9789264037472-en.

———. 2010. *The Economic and Social Role of Internet Intermediaries*. Brussels. http://www.oecd.org/dataoecd/49/4/44949023.pdf.

Oishi, Shigehiro, Ed Diener, and Richard E Lucas. 2007. "The Optimum Level of Well-Being Can People Be Too Happy ?" *Perspectives on Psychological Science* 2 (4): 346–60.

Okulicz-kozaryn, Adam. 2011. "Europeans Work To Live and Americans Live To Work (Who Is Happy to Work More: Americans or Europeans?)." *Journal of Happiness Studies* 12 (2): 225–43. doi:http://dx.doi.org/10.1007/s10902-010-9188-8.

Ostrom, Amy, Mary Jo Bitner, S. W. Brown, K. A. Burkhard, M. Goul, V. Smith-Daniels, H. Demirkan, and E. Rabinovich. 2010. "Moving Forward and Making a Difference: Research Priorities for the Science of Service." *Journal of Service Research* 13 (1): 4–36. doi:10.1177/1094670509357611.

Ott, Myle, Yejin Choi, Claire Cardie, and Jeffrey Hancock. 2011. "Finding Deceptive Opinion Spam by Any Stretch of the Imagination." In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1.*, 309–19.

Ozanne, Julie L, and Laurel Anderson. 2010. "Community Action Research." *Journal of Public Policy & Marketing* 29 (1): 123–37. doi:10.1509/jppm.29.1.123.

Page, Kathryn, and Dianne Vella-Brodrick. 2008. "The 'What', 'Why' and 'How' of Employee Well-Being: A New Model." *Social Indicators Research* 90 (3): 441–58. doi:10.1007/s11205-008-9270-3.

Pak, Alexander, and Patrick Paroubek. 2010. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." In *LREC*, 1320–26.

Pang, Bo, and Lillian Lee. 2005. "Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales." *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* 3 (1): 115–24. doi:10.3115/1219840.1219855.

Paolacci, Gabriele, Jesse Chandler, and Panagiotis Ipeirotis. 2010. "Running Experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5 (5): 411–19.

Pavot, William, and Ed Diener. 1993. "Review of the Satisfaction With Life Scale." *Psychological Assessment* 5 (2): 164–72.

Peffers, Ken, Tuure Tuunanen, Marcus Rothenberger, and Samir Chatterjee. 2007. "A Design Science Research Methodology for Information Systems Research." *Journal of Management Information Systems* 24 (3): 45–77. doi:10.2753/MIS0742-1222240302.

Pennebaker, James. 2013. *The Secret Life of Pronouns: What Our Words Say About Us*. New York, New York, USA: Bloomsbury Press.

Pennebaker, James, Cindy Chung, Molly Ireland, Amy Gonzales, and Roger Booth. 2007. *The Development and Psychometric Properties of LIWC2007*. Austin, TX: University of Texas, Austin.

Pennebaker, James, and Laura King. 1999. "Linguistic Styles: Language Use as an Individual." *Journal of Personality and Social Psychology* 77 (6): 1296–1312.

Pennebaker, James, and Thomas C Lay. 2002. "Language Use and Personality during Crises: Analyses of Mayor Rudolph Giuliani's Press Conferences." *Journal of Research in Personality*. doi:10.1006/jrpe.2002.2349.

Pennebaker, James, Tracy Mayne, and Martha Francsis. 1997. "Linguistic Predictors of Adaptive Bereavement." *Journal of Personality and Social Psychology* 72 (4): 863–71.

Pennebaker, James, Matthias R Mehl, and Kate G Niederhoffer. 2003. "Psychological Aspects of Natural Language Use: Our Words, Our Selves." *Annual Review of Psychology* 54 (January): 547–77. doi:10.1146/annurev.psych.54.101601.145041.

Poczos, Barnabas, Liang Xiong, Dougal J. Sutherland, and Jeff Schneider. 2012. "Nonparametric Kernel Estimators for Image Classification." *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2989–96. doi:10.1109/CVPR.2012.6248028.

Podsakoff, Philip M, Scott B Mackenzie, Jeong-yeon Lee, and Nathan P Podsakoff. 2003. "Common Method Biases in Behavioral Research : A Critical Review of the Literature and Recommended Remedies." *Journal of Applied Psychology* 88 (5): 879–903. doi:10.1037/0021-9010.88.5.879.

Podsakoff, Philip M, Scott B MacKenzie, and Nathan P Podsakoff. 2012. "Sources of Method Bias in Social Science Research and Recommendations on How to Control It." *Annual Review of Psychology* 63 (January): 539–69. doi:10.1146/annurev-psych-120710-100452.

Preziosi, Nadir. 2013. *Life Is Getting Worse in ESS Data: Is This due to Micro or Macro Factors?* 28/2013. Bruges European Economic Research Papers. Bruges.

Procter, Rob, William Housley, Matthew Williams, Adam Edwards, Pete Burnap, Omer Rana, Ewan Klein, et al. 2013. "Enabling Social Media Research Through Citizen Social Science." In *ECSCW 2013 Adjunct Proceedings*, 9–12.

Purdie-Vaughns, Valerie, Claude M Steele, Paul G Davies, Ruth Ditlmann, and Jennifer Randall Crosby. 2008. "Social Identity Contingencies: How Diversity Cues Signal Threat or Safety for African Americans in Mainstream Institutions." *Journal of Personality and Social Psychology* 94 (4): 615–30. doi:10.1037/0022-3514.94.4.615.

Purvis, Alison, Ryan T. Howell, and Ravi Iyer. 2011. "Exploring the Role of Personality in the Relationship between Maximization and Well-Being." *Personality and Individual Differences* 50 (3): 370–75. doi:10.1016/j.paid.2010.10.023.

Qiu, Lin, Han Lin, Angela K Leung, and William Tov. 2012. "Putting Their Best Foot Forward: Emotional Disclosure on Facebook." *Cyberpsychology, Behavior and Social Networking* 15 (10): 569–72. doi:10.1089/cyber.2012.0200.

Radovanovic, Milos, Alexandros Nanopoulos, and Mirjana Ivanovic. 2010. "Hubs in Space: Popular Nearest Neighbors in High-Dimensional Data." *Journal of Machine Learning Research* 11: 2487–2531.

Robbins, Lionel. 1932. *An Essay on the Nature and Significance of Economic Science*. Ludwig von Mises Institute.

Rosenbaum, Mark S, Canan Corus, Amy Ostrom, Raymond P Fisk, Andrew S Gallan, Mario Giraldo, Martin Mende, et al. 2011. “Conceptualisation and Aspriations of Transformative Service Research.” *Journal of Research for Consumers*, no. 19: 1–6.

Ross, Joel, Andrew Zaldivar, Lilly Irani, and Bill Tomlinson. 2010. “Who Are the Turkers ? Worker Demographics in Amazon Mechanical Turk.” In *CHI 2010*, 2863–72.

Rude, Stephanie, Eva-Maria Gortner, and James Pennebaker. 2004. “Language Use of Depressed and Depression-Vulnerable College Students.” *Cognition & Emotion* 18 (8): 1121–33. doi:10.1080/02699930441000030.

Runge, Nina, Niklas Kilian, Jan Smeddinck, and Markus Krause. 2012. “Predicting Crowd-Based Translation Quality with Language-Independent Feature Vectors.” *Workshops at the Twenty- …*, 114–15. http://www.aaai.org/ocs/index.php/WS/AAAIW12/paper/viewPaper/5237.

Russell, Matthew. 2013. *Mining the Social Web*. Second. Sebastopol, CA: O’Reily Media. http://oreilly.com/catalog/errata.csp?isbn=9781449367619 for.

Ruths, Derek, and Jürgen Pfeffer. 2014. “Social Media for Large Studies of Behavior.” *Science* 346 (6213): 1063–64. doi:10.1126/science.346.6213.1063.

Rutledge, R. B., N. Skandali, P. Dayan, and R. J. Dolan. 2014. “A Computational and Neural Model of Momentary Subjective Well-Being.” *Proceedings of the National Academy of Sciences*, August, 1–6. doi:10.1073/pnas.1407535111.

Ryan, Richard M., and Edward L. Deci. 2001. “On Happiness and Human Potentials: A Review of Research on Hedonic and Eudaimonic Well-Being.” *Annual Review of Psychology* 52: 141–66.

———. 2002. “Overview of Self-Determination Theory: An Organismic Dialectical Perspective.” In *Handbook of Self-Determination Research*, 3–33. Rochester, USA: University Rochester Press.

Ryff, Carol D. 1989. “Happiness Is Everything , or Is It ? Explorations on the Meaning of Psychological Well-Being.” *Journal of Personality and Social Psychology* 57 (6): 1069–81.

Ryff, Carol D, and Corey L M Keyes. 1995. “The Structure of Psychological Well-Being Revisited.” *Journal of Personality and Social Psychology* 69 (4): 719–27. http://www.ncbi.nlm.nih.gov/pubmed/7473027.

Ryff, Carol D, and Burton Singer. 2013. “Know Thyself and Become What You Are: A Eudaimonic Approach to Psychological Well Being.” In *The Exploration of Happiness*, edited by Antonella Delle Fave, 97–116. Springer Netherlands.

Saatcioglu, Bige, and Julie L Ozanne. 2013. “Moral Habitus and Status Negotiation in a Marginalized Working-Class Neighborhood.” *Journal of Consumer Research*, July, 000–000. doi:10.1086/671794.

Salas-Zárate, María del Pilar, Estanislao López-López, Rafael Valencia-García, Nathalie Aussenac-gilles, Ángela Almela, and Giner Alor-Hernández. 2014. “A Study on LIWC Categories for Opinion Mining in Spanish Reviews.” *Journal of Information Science* 1 (13): 1–13. doi:10.1177/0165551510000000.

Samman, Emma. 2007. “Psychological and Subjective Well-Being: A Proposal for Internationally Comparable Indicators.” *Oxford Development Studies* 35 (4): 459–86. doi:10.1080/13600810701701939.

Savage, M., and R. Burrows. 2007. “The Coming Crisis of Empirical Sociology.” *Sociology* 41 (5): 885–99. doi:10.1177/0038038507080443.

Schmitt, Manfred, and Martin Do. 1999. “Procedural Injustice at Work, Justice Sensitivity, Job Satisfaction and Psychosomatic Well-Being.” *European Journal of Social Psychology* 29: 443–53.

Scholand, a J, Y R Tausczik, and James Pennebaker. 2010. “Social Language Network Analysis.” *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, 23–26. doi:10.1145/1718918.1718925.

Schwartz, Barry, Andrew Ward, John Monterosso, Sonja Lyubomirsky, Katherine White, and Darrin R. Lehman. 2002. “Maximizing versus Satisficing: Happiness Is a Matter of Choice.” *Journal of Personality and Social Psychology* 83 (5): 1178–97. doi:10.1037//0022-3514.83.5.1178.

Schwartz, H Andrew, Johannes Eichstaedt, Margaret Kern, Lukasz Dziurzynski, Stephanie Ramones, Megha Agrawal, Achal Shah, et al. 2013. “Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach.” *PloS One* 8 (9): e73791. doi:10.1371/journal.pone.0073791.

Schwartz, Norbert, and Gerald Clore. 1983. “Mood, Misinformation, and Judgements of Well-Being: Information and Directive Functions of Affective States.” *Journal of Personality and Social Psychology* 45 (3): 513–23.

Schwartz, Shalom H. 1994. “Are There Universal Aspects in the Structure and Contents of Human Values?” *Journal of Social Issues* 50 (4): 19–45. doi:10.1111/j.1540-4560.1994.tb01196.x.

Scollon, Christie N, C Kim-Prieto, and Ed Diener. 2003. “Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses.” *Journal of Happiness Studies* 4 (1): 5–34.

Seligman, Martin, and Mihaly Csikszentmihalyi. 2000. “Positive Psychology - An Introduction.” *American Psychologist* 55 (1): 5–14.

Sheldon, Kennon M, and Tan H Hoon. 2013. “The Multiple Determination of Well Being : Independent Effects of Positive Traits , Needs , Goals , Selves , Social Supports , and

Cultural Contexts.” In *The Exploration of Happiness*, edited by Antonella Delle Fave, 141–60. Springer Netherlands.

Sheridan-Dodds, Peter, and Christopher M Danforth. 2010. “Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents.” *Journal of Happiness Studies* 11: 441–56. doi:10.1007/s10902-009-9150-9.

Siegel, David. 2012. “The Role of Enticing Design in Usability.” *Interactions* 19: 82–85. doi:10.1145/2212877.2212895.

Skoric, Marko M. 2012. “What Is Slack about Slactivism?” *Methodological and Conceptual Issues in Cyber Activism Research*, 77–104.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Reprinted. London: Deut and Sane.

Smith Warner, Karen. 2013. “The Wellbeing Index : A Landscape of Worldwide Measures and the Potential for Large-Scale Change.” University of Pennsylvania.

Special, Whitney P., and Kirsten T. Li-Barber. 2012. “Self-Disclosure and Student Satisfaction with Facebook.” *Computers in Human Behavior* 28 (2). Elsevier Ltd: 624–30. doi:10.1016/j.chb.2011.11.008.

Spohrer, Jim, and Paul P Maglio. 2010. “Toward a Science of Service Systems.” In *Handbook of Service Science*, 157–94. Springer US.

Stampfl, Nora. 2012. *Die Verspielte Gesellschaft*. Hannover, Germany: Heise Zeitschriften Verlag GmbH & Co KG. http://188.40.159.226/leseproben/3881/1_Inhaltsverzeichnis.pdf.

Steel, Piers, Joseph Schmidt, and Jonas Shultz. 2008. “Refining the Relationship between Personality and Subjective Well-Being.” *Psychological Bulletin* 134 (1): 138–61.

Stevenson, Betsey, and Justin Wolfers. 2008. *Economic Growth and Subjective Well-Being: Reassessing the Easerlin Paradox*. 14282. NBER Working Paper Series. http://www.nber.org/papers/w14282.

Stieglitz, Stefan, and Linh Dang-Xuan. 2012. “Social Media and Political Communication: A Social Media Analytics Framework.” *Social Network Analysis and Mining* 3 (4): 1277–91. doi:10.1007/s13278-012-0079-3.

Stiglitz, Joseph E., Amartya Sen, and Jean-Paul Fitoussi. 2009. “Report by the Commission on the Measurement of Economic Performance and Social Progress.” *SSRN Electronic Journal*. doi:10.2139/ssrn.1714428.

Stone, Lori, and James Pennebaker. 2002. “Trauma in Real Time: Talking and Avoiding Online Conversations About the Death of Princess Diana.” *Basic and Applied Psychology* 24 (3): 173–83.

Tausczik, Yla, and James Pennebaker. 2010. “The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods.” *Journal of Language and Social Psychology* 29 (1): 24–54. doi:10.1177/0261927X09351676.

Taylor, L., Ralph Schroeder, and E. Meyer. 2014. “Emerging Practices and Perspectives on Big Data Analysis in Economics: Bigger and Better or More of the Same?” *Big Data & Society* 1 (2). doi:10.1177/2053951714536877.

Thelwall, Mike, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. “Sentiment Strength Detection in Short Informal Text.” *Journal of the American Society for Information Science and Technology* 61 (12): 2544–58. doi:10.1002/asi.21416.

Thinley, Jigmi. 2011. “Gross National Happiness: A Holistic Paradigm for Sustainable Development.” India: Bhutan.

Tinati, R., S. Halford, L. Carr, and C. Pope. 2014. “Big Data: Methodological Challenges and Approaches for Sociological Analysis.” *Sociology* 48 (4): 663–81. doi:10.1177/0038038513511561.

Tov, William, Kok Leong Ng, Han Lin, and Lin Qiu. 2013. “Detecting Well-Being via Computerized Content Analysis of Brief Diary Entries.” *Psychological Assessment* 25: 1069–78.

Tsur, Oren, and Ari Rappoport. 2010. “ICWSM – A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews.” *ICWSM*, 1–9.

Tumasjan, Andranik, Timm O Sprenger, Philipp G Sandner, and Isabell M Welpe. 2010. “Predicting Elections with Twitter : What 140 Characters Reveal about Political Sentiment.” *ICWSM* 10: 178–85.

Turney, Peter D, and Patrick Pantel. 2010. “From Frequency to Meaning: Vector Space Models of Semantics.” *Journal of Artificial Intelligence Research* 37: 141–88.

Utz, Sonja. 2005. “Types of Deception and Underlying Motivation: What People Think.” *Social Science Computer Review* 23 (1): 49–56. doi:10.1177/0894439304271534.

Utz, Sonja, Martin Tanis, and Ivar Vermeulen. 2012. “It Is All about Being Popular: The Effects of Need for Popularity on Social Network Site Use.” *Cyberpsychology, Behavior and Social Networking* 15 (1): 37–42. doi:10.1089/cyber.2010.0651.

Vaillant, George. 2008. *Aging Well: Surprising Guideposts to a Happier Life From the Landmark Harvard Study of Adult Development*. New York, New York, USA: Hachette Book Group. doi:10.1176/appi.ajp.161.1.178.

Van Suntum, Ulrich. 2012. *Zur Kritik Des BIP Als Indikator Für Wohlstand Und Wirtschaftswachstum*. Vol. 17. Münster, Germany. http://www.bdi.eu/images_content/KonjunkturStandortUndWettbewerb/BIP-Studie_im_Auftrag_des_BDI_final_s-w.pdf.

Varelius, Jukka. 2013. “Objective Explanations of Individual Well-Being.” In *The Exploration of Happiness*, edited by Antonella Delle Fave, 15–30. Springer Netherlands.

Vargo, Stephen L, Paul P Maglio, and Melissa Archpru Akaka. 2008. “On Value and Value Co-Creation: A Service Systems and Service Logic Perspective.” *European Management Journal* 26 (3): 145–52. doi:10.1016/j.emj.2008.04.003.

Vargo, Stephen L. 2009. “Toward a Transcending Conceptualization of Relationship: A Service-Dominant Logic Perspective.” Edited by Jaqueline Pels. *Journal of Business & Industrial Marketing* 24 (5/6): 373–79. doi:10.1108/08858620910966255.

Vargo, Stephen L., and Melissa Archpru Akaka. 2009. “Service-Dominant Logic as a Foundation for Service Science: Clarifications.” *Service Science* 1 (1): 32–41.

Vargo, Stephen L., and Robert F. Lusch. 2008. “From Goods to Service(s): Divergences and Convergences of Logics.” *Industrial Marketing Management* 37 (3): 254–59. doi:10.1016/j.indmarman.2007.07.004.

Vassileva, Julita. 2012. “Motivating Participation in Social Computing Applications: A User Modeling Perspective.” *User Modeling and User-Adapted Interaction* 22: 177–201. doi:10.1007/s11257-011-9109-5.

Veenhoven, Ruut. 1984. “Conditions of Happiness.” Erasmus University.

———. 2008. *Measures of Gross National Happiness*. Istanbul.

———. 2010. “Greater Happiness for a Greater Number.” *Journal of Happiness Studies* 11 (5): 605–29. doi:10.1007/s10902-010-9204-z.

———. 2013. “The Four Qualities of Life Ordering Concepts and Measures of the Good Life.” In *The Exploration of Happiness*, edited by Antonella Delle Fave, 195–226. Springer Netherlands.

Vella, Kellie, and Daniel Johnson. 2012. “Flourishing and Video Games.” In *Proceedings of The 8th Australasian Conference on Interactive Entertainment Playing the System - IE ’12*, 1–3. New York, New York, USA: ACM Press. doi:10.1145/2336727.2336746.

Vella, Kellie, Daniel Johnson, and Leanne Hides. 2013. “Positively Playful: When Videogames Lead to Player Wellbeing.” *First International Conference on Gameful Design, Research and Applications*, 99–102. doi:10.1145/2583008.2583024.

Wang, N, Michal Kosinski, David Stillwell, and J Rust. 2014. “Can Well-Being Be Measured Using Facebook Status Updates ? Validation of Facebook ’ S Gross National.” *Social Indicators Research* 115 (1): 483–91. doi:10.1007/s11205-012-9996-9.

Waterman, Alan S. 1990. “The Relevance of Aristotle’s Conception of Eudaimonia for the Psychological Study of Happiness.” *Theor. & Philo. Psych.* 10 (1): 39–44.

———. 1993. “Two Conceptions of Happiness: Contrasts of Personal Expressiveness (eudaimonia) and Hedonic Enjoyment.” *Journal of Personality and Social Psychology* 64 (4): 678–91. doi:10.1037//0022-3514.64.4.678.

———. 2007. “On the Importance of Distinguishing Hedonia and Eudaimonia When Contemplating the Hedonic Treadmill.” *The American Psychologist* 62 (6): 612–13. doi:10.1037/0003-066X62.6.612.

Waterman, Alan S., Seth J. Schwartz, and Regina Conti. 2006. “The Implications of Two Conceptions of Happiness (Hedonic Enjoyment and Eudaimonia) for the Understanding of Intrinsic Motivation.” *Journal of Happiness Studies* 9 (1): 41–79. doi:10.1007/s10902-006-9020-7.

White, Sarah, and Jethro Pettit. 2004. *Participatory Approaches and the Measurement of Human Well-Being*. 2004/57. Helsinki.

Wilckens, Max, and Margeret Hall. 2015. “Can Well-Being Be Predicted ? A Machine Learning Approach.” *SSRN Electronic Journal*. doi:http://dx.doi.org/10.2139/ssrn.2562051.

Wilson, Robert E, Samuel D Gosling, and Lindsay T Graham. 2012. “A Review of Facebook Research in the Social Sciences.” *Perspectives on Psychological Science* 7 (3): 203–20. doi:10.1177/1745691612442904.

Wilson, Theresa, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. “OpinionFinder : A System for Subjectivity Analysis.” In *Proceedings of HLT/EMNLP on Interactive Demonstrations*, 34–35. Association for Computational Linguistics.

Wilson, Warner. 1967. “Correlates of Avowed Happiness.” *Psychological Bulletin* 67 (4): 294–306. doi:10.1037/h0024431.

Winter, Robert. 2008. “Design Science Research in Europe.” *European Journal of Information Systems* 17: 470–75. doi:10.1057/ejis.2008.44.

Wolf, Markus, Andrea Horn, Matthias Mehl, Severin Haug, James Pennebaker, and Hans Kordy. 2008. “Computergestützte Quantitative Textanalyse.” *Diagnostica* 54 (2): 85–98. doi:10.1026/0012-1924.54.2.85.

Wu, Chia-Huei, and Grace You. 2007. “Examining the Relationship between Global and Domain Measures of Quality of Life by Three Factor Structure Models.” *Social Indicators Research* 84: 189–202. doi:10.1007/s11205-006-9082-2.

Yang, Hongwei. 2013. “The Case for Being Automatic: Introducing the Automatic Linear Modeling (LINEAR) Procedure in SPSS Statistics.” *Multiple Linear Regression Viewpoints* 39 (2): 27–37.

Yarkoni, Tal. 2010. “Personality in 100,000 Words: A Large-Scale Analysis of Personality and Word Use among Bloggers.” *Journal of Research in Personality* 44 (3): 363–73.

Youyou, Wu, Michal Kosinski, and David Stillwell. 2015. “Computer-Based Personality Judgments Are More Accurate than Those Made by Humans.” *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1418680112.

Zamagni, Stefano. 2014. “Public Happiness in Today’s Economics.” *International Review of Economics*, 1–8. doi:10.1007/s12232-014-0209-5 Research.

Zhang, Ping. 1993. “Model Selection Via Multifold Cross Validation.” *The Annals of Statistics* 21 (1): 299–313.

Zhao, Shanyang, Sherri Grasmuck, and Jason Martin. 2008. “Identity Construction on Facebook: Digital Empowerment in Anchored Relationships.” *Computers in Human Behavior* 24 (5): 1816–36. doi:10.1016/j.chb.2008.02.012.

Zhong, F, Steven O. Kimbrough, and Dj Wu. 2002. “Cooperative Agent Systems: Artificial Agents Play the Ultimatum Game.” In *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*, 00:2207–15. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=994150.

Zimmer, Michael. 2010. “‘But the Data Is Already Public’: On the Ethics of Research in Facebook.” *Ethics and Information Technology* 12 (4): 313–25. doi:10.1007/s10676-010-9227-5.