This is a repository copy of *Affect dynamics in data science practice*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/151247/

Version: Accepted Version

Proceedings Paper:

Affect Dynamics in Data Science Practice

Jo Bates
Information School
The University of Sheffield
Sheffield, UK
jo.bates@sheffield.ac.uk

David Cameron
Information School
The University of Sheffield
Sheffield, UK
d.s.cameron@sheffield.ac.uk

Abstract

While data science is often imagined as a dispassionate and objective process of extracting information from data, ethnographic work has identified that affect and emotion play an important role in shaping data science practice. In this paper, we consider how HCI/psychology approaches might contribute to the further development of these exploratory findings, and inform the development of methods to enhance data scientists' awareness of when affect dynamics may be influencing their practice.

Author Keywords

Affect; Emotion; Data Science; Data Science Practice

ACM Classification Keywords

H.5.0 [HCI]: General; K.7.0 [Social and Professional Topics]: General

Introduction

In what ways do affect and emotion shape people's engagement in Data Science tasks? How might these affective engagements impact upon results and outputs? How can we help Data Scientists to become more aware of how affective dynamics might be influencing their practice?

Within the popular imaginary, data science is often framed as a process in which the scientist engages in the dispas-

sionate pursuit of an objective result or outcome. However, this model asks much of a data scientist. Even ignoring the potential for the research topic to be an emotive one for the data scientist, the process of pursuing a goal-driven activity is regarded within the psychology literature as essentially an emotional one [2].

There is currently little understanding of the affective dynamics of data scientists' engagements with their practice, nor how these dynamics may impact upon the outputs produced and the wider social implications. Across various research and activist communities there is a growing demand to address issues of justice, reflexivity and context in practices of data handling, data analysis and visualisation, and machine learning and AI. For example, at conferences such as Data Justice 2018, Data Power 2015/17/19, ACM FAT* and Data4blacklives, and in the work of academics [4, 5, 10].

We argue that if we are to influence the development of data science as a just and socially beneficial practice, an important component of this work will be to foster critical and reflexive practices within the Data Science community. To do this, a much deeper understanding of the affective and emotional dynamics at play when data scientists do their work is required, as are ways to operationalise these findings within the Data Science practitioner context.

The nature of the relationship between data and emotion has rarely been examined [6]. Only the work of [9] and [6] draws attention to the importance of understanding the emotional dynamics of how people make sense of data. However, as Kitchin identifies, subjectivities, and thus feelings, emotions and affects, are a key component in the complex data assemblages that frame "what is possible, desirable and expected of data" [7](p. 24).

Pilot ethnographic research identified that feeling, emotion and affect play an important role in how Data Scientists engage with the inputs, outputs and practices of a simple machine learning task [1]. Through participant observation and interviews with members of a small data science project team, they observed complex affective dynamics in relation to every stage of the data mining process: data acquisition, data cleaning, building a classifier, and the numerical and visual outputs. They also observed that the ways in which feeling and emotion framed participants' understanding and empathy with the people behind the data - the data subjects.

Importantly, they also observed that these emotional utterances were only made explicitly in individual interviews. In project team meetings, email communications and project reports these affective dynamics were hidden from team members, suggesting teams were often unaware of the emotional undercurrents shaping participants' engagements with the project. The research findings suggested that these affective dynamics may have impacted upon crucial decisions, including a decision to oversample to achieve a high, but ultimately for the end-users implausible, level of predictive accuracy for their model.

While these early findings point to a key area for further investigation, the findings of a small-scale pilot ethnographic study such as [1] are challenging to translate for, and achieve credibility within, much of the Data Science community. Our research direction aims to build on this initial work in an interdisciplinary collaboration between Critical Data Studies (cultural/information studies) and HCI (psychology).

From an HCI/psychology perspective, we may approach this topic in terms of the emotional dynamics in goal pursuit [2]. Goals may be abstract/concrete and distal/immediate; they may interlink so that single actions serve multiple

goals, or compete so progress towards one goal impedes progress to another [8]. The Control Theory Framework maps out a hierarchy by which goals may arrange to form structured and meaningful behaviour in individuals [3]. An example hierarchy that may begin to explain the affective dynamics of data science practice is shown in Table 1.

Demonstrate effective Data Scientist profile	↓ How? ↓
Complete the current project well	
Develop a strong model	
Improve data fit	
Add/remove variables	
Use data analysis interface	
Extend arm and click mouse button	↑ Why? ↑

Table 1: Possible hierarchy of goals for a data scientist

At each point, we can examine *how* one expects to achieve their goal and move down the hierarchy. Conversely, at each point we can examine *why* and move up the hierarchy. Ultimately, even the grandest goals require low-level actions to progress. Goal-directed behaviour is regarded to directly impact experienced affect [2] - affect indicates the rate of progression towards a goal. Positive progress towards a goal (or away from an aversive goal) is experienced as positive affect and the reverse as negative affect [2].

This structure may explain the affect experienced by the DS researchers in the prior ethnographic study [1]:

- Progress towards the goal *complete the current project well* is uncertain: there are many ways for this to occur, perhaps it wasn't clear what that would look like.

- The goal of *develop a strong model* may be viewed as having a particular process (e.,g *improve data fit*)

- Attention may be focused on lower goals where more positive progress can be seen. What constitutes a strong model is abstract and subject to others' evaluation, whereas positive progress towards *improve data-fit* is concrete and inarguable: one number is bigger than the other.

- The process of data preparation is described as being painful. The value of data preparation may not be appreciated, or potentially, this activity may be seen as interfering with progress towards other goals.

- The higher-tier goals are not attended to, progress on *complete the current project well* is not monitored as DS researchers get off track. Alternatively, DS researchers misinterpret the *improve data fit* goal as being a direct route to achieve the higher goal.

We envisage that through such an integration of Critical Data Studies and HCI approaches to the study of the affective dynamics of Data Science practices, we might also be able to develop new types of recommendations that enhance practice. For example, beyond integrating emotional awareness into the teaching of critical and reflective Data Science practice, we hope that integrating methods and techniques from HCI will allow us to develop and test different forms of guidance and alerts that may help enhance data scientists' awareness of when affective dynamics may be influencing their practice, for example:

- Education of Data Scientists on appropriate monitoring of project goals e.g. Data cleaning is an essential part of the progress towards a goal.

- Locking out behaviours for Data Scientists until early work is completed i.e. to address the temptation to

analyse data ahead of it being ready / frustration that tools (toys) to use are in plain sight.

- Effective management that is able to identify that Data Scientists' goals are off track.

- Self-monitoring notifications to nudge attention towards higher goal of *develop a strong model* e.g. Warnings on overfitting, require users specify a target for each session, interface reminders that the user has tried [x] model variants in a particular session.

Authors' Interest

Jo Bates is Senior lecturer in Information Politics and Policy at the University of Sheffield's Information School. She has been actively engaged in the Critical Data Studies community since its inception. She is a co-founder/co-organiser of the Data Power conference, and on the editorial board of Big Data and Society journal. She has published in the fields of Critical Data Studies and Information Studies, as well as contributing to interdisciplinary publications with Information and Computer Scientists. She is Deputy Programme Co-ordinator for the MSc Data Science at University of Sheffield, and runs the core Data and Society module for this degree. She was PI on the pilot project reported in [1], and is interested in exploring the potential for incorporating an HCI approach to take this work forward.

Dave Cameron is a Lecturer in Human-Computer Interaction at the University of Sheffield's Information School and coordinates the department's module for User Interface Design and Human- Computer Interaction. Having trained as a social psychologist, he has a research interest in the impact affect has on goal-directed behaviour. His PhD explored the recurrent influence affect and affect regulation dynamics have on goal pursuit and the process of switching focus from one goal to another.

REFERENCES

1. Jo Bates and Jess Elmore. 2018. Identifying the Affective Dimension of Data Mining Practice: An Exploratory Study. In *Transforming Digital Worlds*, Gobinda Chowdhury, Julie McLeod, Val Gillet, and Peter Willett (Eds.). Springer International Publishing, Cham, 243–252.

2. Charles Carver and Michael Scheier. 1990. *Principles of self-regulation: Action and emotion*. Guilford Press.

3. Charles Carver and Michael Scheier. 1998. *On the Self-regulation of Behavior*. Cambridge University Press.

4. Lina Dencik, Arne Hintz, and Jonathan Cable. 2016. Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society* 3, 2 (2016), 2053951716679678.

5. Catherine D'Ignazio and Lauren F Klein. In preparation. *Data Feminism*. MIT Press Open.

6. Helen Kennedy and Rosemary Lucy Hill. 2018. The feeling of numbers: Emotions in everyday engagements with data and their visualisation. *Sociology* 52, 4 (2018), 830–848.

7. Rob Kitchin. 2014. *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

8. William T Powers. 1973. *Behavior: The control of perception*. Aldine Chicago.

9. Minna Ruckenstein. 2014. Visualized and interacted life: Personal analytics and engagements with data doubles. *Societies* 4, 1 (2014), 68–84.

10. Linnet Taylor. 2017. What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society* 4, 2 (2017), 2053951717736335.