Editorial to the Special Edition of the Journal of Legal Research Methodology on 'Empirical Legal Research'

Malvika Unnithan, Lyndsey Bengtsson, Paul Dargue, Anqi Shen

Following the successful inaugural edition of the *Journal of Legal Research Methodology*, our second volume is a special edition focused on 'empirical legal research methodology'. We use this term broadly to encompass qualitative, quantitative, and mixed methods involving the collection or creation of new data as part of the study of law, legal processes and legal phenomena. It has been widely noted that empirical legal research allows for exploration of the law world beyond its legal rules. The use of social sciences research methods has been known to allow empirical questions to be answered in legal studies, especially in relation to how the law is understood and used to make decisions in practice. This, we think, gives empirical studies a unique and important place in legal research to investigate and create a meaningful impact on the function of law in society.

Despite being such a crucial aspect of legal study, it has been observed that many students and early career academics carrying out empirical legal research come from academic backgrounds which are traditionally focused primarily on doctrinal legal research, resulting in limited exposure to social research methods. As a result, legal researchers start out having engaged predominantly with the findings of empirical legal research, rather than being encouraged to consider methodological issues. With this edition of the journal, we aimed to give authors the opportunity to reflect on the research processes employed in their study to enable readers to judge how the research data may be used. We invited critical discussions on the practicalities of the methodologies employed regarding issues such as, access to participants, the strength and weakness of the methodology used, and the reliability, validity, and representativeness of the data obtained to stress research rigour. Our call for papers resulted in four thought-provoking contributions.

This edition begins with an article entitled '*Access to Justice Software Development, Participatory Action Research Methods and Researching the Lived Experiences of British Military Veterans*'. Olusanya et al, reflect on their experiences of developing the UK's first access to justice platform for veterans and their families through an ongoing Participatory Action Research (PAR) project. In this article, they present findings from their 3-Stage research process brought about through their work with armed forces veterans, representatives from veterans' service providers, and the Veterans Legal Link team members comprising of legal academics, lawyers, sociologists, computer software designers and graphic designers, in order to address issues related to the delivery of access to justice. Their aim with this piece is to contribute to the limited but growing literature on PAR in the field or law, and to also demonstrate the ways in which PAR methodology can be useful to access to justice research projects. This article provides pragmatic insight into the benefits and challenges of engaging in a sustained PAR project, whilst also advocating for the use of this methodology in research focused on investigating and solving social problems where a gap between theory and practice exists.

The next article in this edition by Pina-Sanchez and Gosling, *'Enhancing the Measurement of Sentence Severity through Expert Knowledge Elicitation'*, contributes to both measurement and sentencing literature in three main ways as outlined by the authors. They did so firstly, by testing a key assumption made in studies estimating the relative severity of different sentence types, secondly, by noting the wide differences in the range of severity covered by some of the main disposal types used in England and Wales elicited from six sentencing experts, and thirdly by presenting a new scale of sentence severity through a modified version of the Thurstone method which allows for unequal variances. This article highlights the challenges with the assumption of equal variances in the standard Thurstone scaling method and demonstrates how it can be relaxed using data collected from expert knowledge elicitation techniques. The research is said to have resulted in a proposed new scale of severity which can be used as an analytical tool to help facilitate more robust and quantitative sentencing research.

Redhead's article *'From Legislative Intent to Hospice Practice: Exploring the Genealogy of The Mental Capacity Act 2005'*, provides insight into the 'life story' of the Act and how it is understood and interpreted in practice. Redhead takes the reader through the four distinct yet linked phases of the research process, starting with a description of the qualitative methods developed and used to trace the key ideas of the policy-makers and legislators during the formation of the Mental Capacity Act 2005 (legislative intent), all the way through to, current practice based on the perceptions on the law held by professionals in hospices. The findings presented in this article focus on the patient's role in the decision-making process in cases where they lack capacity. The discussion and reflections in this article provide a strong case for the use of a Foucauldian genealogical approach along with a phased combination of documentary and empirical enquiry when investigating the 'life story' of any statute.

To round up this edition, Bleazby's article, *'Take (what they say) with a pinch of salt: Engaging in Empirical Research to Understand the Parameters of the 'Quality' in 'Poor-Quality Defence Lawyering'* draws from the author's PhD thesis which discusses the quality of defence legal assistance and attempts to proffer a common definition or standard of the term 'quality' in this context. This article focuses on the empirical data acquired from semi-structured interviews held with defence lawyers on their perceptions, opinions and experiences of 'quality' in defence representation. It highlights that the law is a social construction that cannot be advanced in isolation from its interpretation and application, and thus puts forth an argument for developing, articulating and testing legal theory through empirical research methodologies.

Each of the articles presented in this special edition, provide valuable insights into the practicalities of empirical legal research in a range of different legal contexts. Through their experiences, the authors provide reflective and pragmatic advice for researchers considering or undertaking empirical legal research. These well-argued articles make important contributions to legal research and the academic communities that engage with it. We congratulate the authors on their research success.

Access to Justice software development, Participatory Action Research Methods and Researching the Lived Experiences of British Military Veterans

Olaoluwa Olusanya, William George Andrew Collier, Simon Marshall, Victoria Knapp and Alex Baldwin

Keywords: Access to justice; participatory action research; military veterans

Abstract

Participatory action research (PAR) methods aim to position the people who are most affected by the issue being studied as equal partners in the research process through a cyclical process of data gathering, data analysis, planning and implementing action and evaluation and reflection. In doing so, it ensures that the research better reflects participants' ideas, priorities, and needs, thereby enhancing its validity and relevance and the support for the findings and proposed changes. Furthermore, it generates immediately applicable results. In this paper, we reflect on our experiences of developing the UK's first access to justice platform for veterans and their families through an ongoing PAR project that brought together armed forces veterans, representatives from veterans' service providers, and the Veterans Legal Link team members comprising of legal academics, lawyers, sociologists, computer software designers and graphic designers to collect, interpret, and apply community information to address issues related to the delivery of access to justice. We present findings from Stages 1 and 2 of our three-stage iterative research process which includes the following steps: Understanding and cross-checking the lived experience of the veteran community (Stage 1), developing and testing a prototype of the access to justice platform (Stage 2) and creating the final product and giving real users an opportunity to use the platform (Stage 3). Data collection and analysis from Stage 1 of the study informed the themes that underpinned Stage 2. Specifically, data was collected through the following methods: co-facilitated focus group discussions, a web survey that was codesigned with veteran community stakeholders and remote and digitally enabled ethnographic research methods. We include several reflections that may help legal practitioners and researchers interested in applying PAR within the area of access to justice and the field of legal research.

1. Introduction

As a form of applied research, participatory action research (PAR)[1] has been used extensively in many disciplines however it is comparatively rare in the field of law[2] and even less common in the area of access to justice research.[3] In this paper, we reflect on our experiences of developing the UK's first access to justice platform for veterans (former members of the British Armed forces who served for at least one day[4]) and their families through an ongoing PAR project that brought together armed forces veterans, representatives from veterans' service providers, and the Veterans Legal Link (VLL) team members comprising of legal academics, lawyers, sociologists, computer software designers and graphic designers to collect, interpret, and apply community information to address issues related to the delivery of access to justice. In doing so, our aims were to contribute to the small but growing literature on PAR in the field of law and to demonstrate the usefulness of PAR methodology to access to justice research projects. This Article proceeds as follows. Section 2 introduces the wider VLL project which had a catalytic and spin-off effect on the access to justice platform project thereby serving as a prelude to the subsequent sections. In Section 3, we unpack the methodological and epistemological foundations of PAR, discuss the origins of PAR and compare PAR and

[1] For relevant works see WF Whyte, DJ Greenwood and P Lazes, 'Participatory action research: Through practice to science in social research' in WF Whyte (ed.), *Participatory Action Research* (Sage 1991) 19; AE Brodsky and EA Welsh, 'Applied Research' in LM Given (ed.), *The Sage Encyclopedia of Qualitative Research Methods* (Sage 2008) Vol. 1, 17.

[2] See PJ Bentley, M Gulbrandsen and S Kyvik, 'The relationship between basic and applied research in universities' (2015) 70(4) Higher Education 689, 700; for relevant examples see EMS Houh and K Kalsen 'It's critical: Legal participatory action research' (2014) 19 Michigan Journal of Race & Law, 287; J Moore, M Sandys and R Jayadev, 'Make Them Hear You: Participatory Defense and the Struggle for Criminal Justice Reform' (2014) 78 Albany Law Review 1281; E Rosario-Moore and A Rosario-Moore, 'From the Ground Up: Criminal Law Education for Communities Most Affected by Mass Incarceration' (2016) 23 Clinical Law Review 753.; AA Akbar, SM Ashar and J Simonson, 'Movement Law' (2021) 73(4)Stanford Law Review 821.

[3] See e.g. Y Maker, J Offergeld and A Arstein-Kerslake, 'Disability Human Rights Clinics as a model for teaching Participatory International Human Rights Lawyering' (2018) Int'l J. Clinical Legal Educ. 23, 46; MA Moss, 'The Escambia Project: An Experiment in Community-Designed Justice' (2020) 36 (3) Design Issues, 45.

[4] See e.g. H Burdett, C Woodhead, AC Iversen, S Wessely, C Dandeker, NT Fear '"Are you a veteran?" Understanding of the term "veteran" among UK ex-service personnel: A research note' (2013) 39(4) Armed Forces & Society 752.

conventional research[5] thereby providing the groundwork for Section 4. In Section 4, we present findings from Stages 1 and 2 of our three-stage iterative research process which incorporates the following steps: Understanding and cross-checking the lived experience of the veteran community (Stage 1), developing and testing a prototype of the access to justice platform (Stage 2) and creating the final product and giving users an opportunity to use the platform (Stage 3). As elaborated in more detail below, data collection and analysis from Stage 1 of the study informed the themes that underpinned Stage 2. Specifically, data was collected through the following methods: co-facilitated focus group discussions, a web survey that was codesigned with veteran community stakeholders and remote and digitally enabled ethnographic research methods. Finally, in Section 5, we discuss the benefits and challenges of PAR, while also offering several reflections that may help legal practitioners and researchers interested in applying PAR within the area of access to justice and the field of legal research.

2. The Veterans Legal Link project

The Veterans Legal Link (VLL) project is an access to justice project that provides free legal advice services and professional signposting for veterans and their families. The VLL's services were accessible through the use of drop-in centres across Wales, as well as being accessible via phone and email. The VLL's services are available to any British military or blue-light veterans and their families. The VLL grew out of the Principal Investigator's research interests and was established in 2015 in response to the introduction of the Legal Aid, Sentencing and Punishment of Offenders Act 2012 (LASPO). LASPO was introduced as part of the government's programme of spending cuts to achieve significant savings to the legal aid

[5] For a detailed comparison between PAR and conventional research see A Cornwall and R Jewkes 'What is participatory research?' (1995) 41(12) Social Science Methodology 1667. See also R Chambers 'The Origins and Practice of Participatory Rural Appraisal' (1994) 22 (7) World Development, 953.

budget. However, LASPO has had a disproportionate impact on veterans and other vulnerable populations. For instance, the Amnesty International report "Cuts that Hurt"[6] observed that LASPO has had an unequal impact on people with additional vulnerabilities and or disadvantages that make accessing, navigating and understanding the legal process harder. This includes those with mental illnesses, low numeracy and literacy levels, and alcohol and drug conditions.[7]

The VLL project had been in existence for six years—the minimum recommended number for a long-standing academic-community partnership[8]. In 2019, we conducted a usability evaluation (as there was a continuing trend of our service users accessing our service through routes other than the drop-in centres) to better understand VLL's service users and their needs and discovered that only 30% of the service users were accessing the service via the drop-in clinics. This presented the VLL with an opportunity to begin research into the optimal mode of delivery for the service on an ongoing basis. This started our journey using PAR as our methodology in order to, "make sure the questions asked and methods used do justice to the pressing issues at hand, the richness of participant knowledge and local views about the matters under investigation."[9] The first phase study involved understanding and cross-checking the lived experience of the veteran community. We carried out an initial consultation in July 2019 with multiple veteran organisations and the wider veterans' community to better understand their lived experience when accessing services and the wider needs of the community (addressed in greater detail in section 4 on the research process). The initial consultation indicated a need for an additional complementary online provision for the delivery of the

[6] Amnesty International, *Cuts that hurt: the impact of legal aid cuts in England on access to justice* (2016). Available at: https://www.amnesty.org.uk/files/aiuk_legal_aid_report.pdf
[7] ibid 4.
[8] BL Brush et al. 'Success in Long-Standing Community-Based Participatory Research (CBPR) Partnerships: A Scoping Literature Review' (2020) 47(4) Health education & behavior 556.
[9] JM Chevalier, M Jacques and DJ Buckles. *Participatory Action Research: Theory and Methods for Engaged Inquiry* (Taylor & Francis Group 2013), 5.

VLL's services. This crystallized into the access to justice platform project. Phase 2 involved the creation of an online mobile and web-based portal for the delivery of the service and was greeted with enthusiasm and positivity by the veterans and organisations consulted in the initial stage.

The access to justice platform project was led by the VLL and guided by a steering group composed of armed forces veterans, representatives from veterans' service providers (both public and private) and veteran organisations. The Veterans Legal Link (VLL) team was comprised of legal academics, lawyers, sociologists, computer software designers and graphic designers and aimed to collect, interpret, and apply community information to address issues related to the delivery of access to justice. From its very beginnings, the research methodology into the veterans, community was firmly rooted in the PAR approach.

3. Participatory Action Research (PAR)

PAR encapsulates an epistemological position[10], a research methodology,[11] and a process for collaborative social action.[12] Emerging from Lewin's development of the Action Research methodology in the 1940s and 1950s and influenced by several intellectual traditions including interpretivism/constructivism[13] and critical theories[14], PAR aims to challenge power dynamics in conventional research methods by critically assessing the researcher-researched relationship

[10] See P Freire, *Pedagogy of the oppressed* (Bloomsbury 2018); O Fals Borda, *Knowledge and people's power: Lessons with peasants in Nicaragua, Mexico and Colombia* (New Horizons Press 1988).
[11] See C MacDonald, 'Understanding participatory action research: A qualitative research methodology option' (2012) 13(2) The Canadian Journal of Action Research 34.
[12] See TE Benjamin-Thomas, AM Corrado, C McGrath, DL Rudman and C Hand, 'Working Towards the Promise of Participatory Action Research: Learning From Ageing Research Exemplars' (2018) International Journal of Qualitative Methods https://doi.org/10.1177/1609406918817953.
[13] See IP Canlas and M Karpudewan, 'Blending the Principles of Participatory Action Research Approach and Elements of Grounded Theory in a Disaster Risk Reduction Education Case Study' (2020) International Journal of Qualitative Methods https://doi.org/10.1177/1609406920958964; PJ Kelly, 'Practical suggestions for community interventions using participatory action research' (2005) 22 (1) Public health nursing 65.
[14] See F Baum, C MacDougall and D Smith, 'Participatory action research' (2006) 60 (10) Journal of epidemiology and community health 854.

and valuing lived experience and local knowledge[15]. A wide variety of definitions of PAR have been proposed, but the definition suggested by the Institute of Development Studies most closely resonates with our understanding of PAR:

> PAR focuses on social change that promotes democracy and challenges inequality; is context-specific, often targeted on the needs of a particular group; is an iterative cycle of research, action and reflection; and often seeks to 'liberate' participants to have a greater awareness of their situation in order to take action".[16]

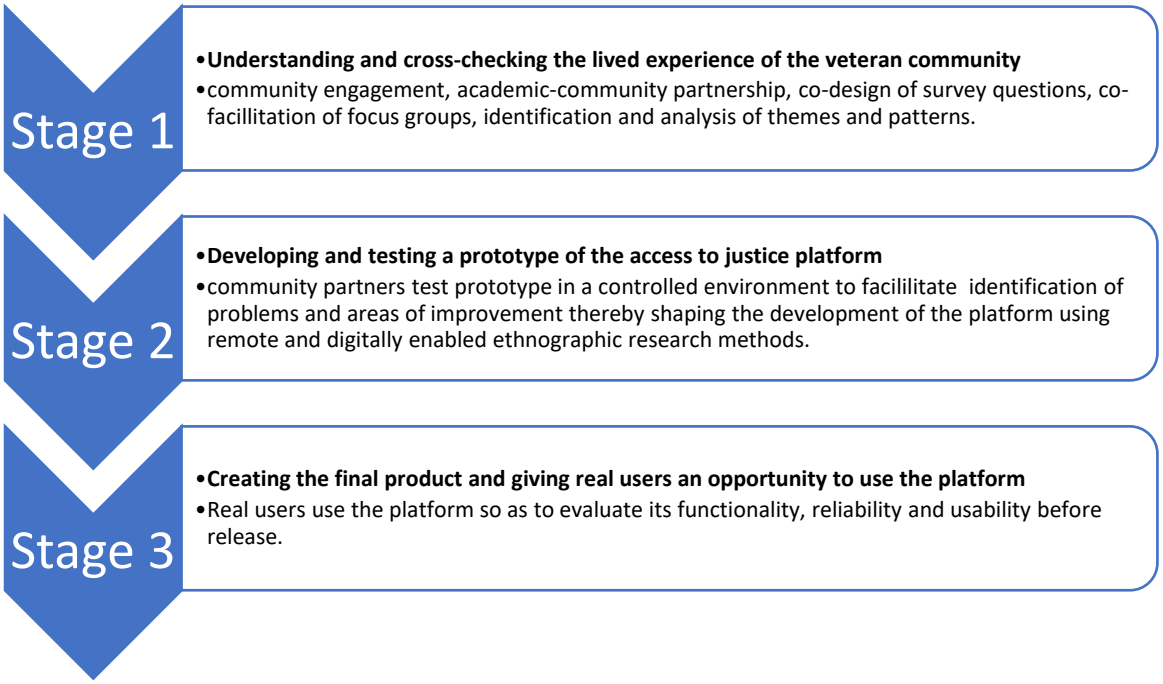
One of the aims of our study was to challenge the inequality in access to justice for veterans and their families living on a low income in rural and remote communities of the UK. We employed an iterative methodology[17] underpinned by a continuous cycle of data gathering, data analysis, formulation of an action plan, implementation of the action plan and evaluation (Figure 1).

Figure 1: Action research cycle

[15] See R Chambers, *Rural apprasial: rapid, relaxed and participatory* (Institute of Development Studies (UK) 1992) 311.
[16] Institute of Development Studies, 'Glossary' (2018) Participatory Methods (Online), , https://www.participatorymethods.org/glossary/participatory-action-research [Accessed 25 January 2022]
[17] See e.g. S Kemmis and R McTaggart, *The Action Research Planner* (Deakin University Press.1988).

Moreover, through the project steering group, composed of academic researchers and community partners, our PAR approach incorporated continuous dialogue, relationship building and active and genuine participation of veteran community stakeholders in the research process and thus had a liberating effect on them.[18] In addition, as "there is no one way to implement PAR"[19] we elected to use co-facilitated focus group discussions, a web survey that was codesigned with veteran community stakeholders and remote and digitally enabled ethnographic research methods to undertake our research at the appropriate phases (outlined in Figure 2). With the onset of Covid-19 it was particularly important to use a digitally enabled method for phase 2.

Figure 2: A three-stage iterative process for developing the access to justice platform



[18] K Lewin, 'Action research and minority problems' (1946) 46 J. Soc. Issues 34; P Freire, *Pedagogy of the Oppressed* (1968. Trans. Myra Bergman Ramos. New York: Herder 1972).
[19] GW White, M Suchowierska and M Campbell, 'Developing and systematically implementing participatory action research' (2004) 85 Archives of physical medicine and rehabilitation, 3.

Drawing on the experiences of researchers from other disciplines, it is evident that PAR as a method of research has both benefits and challenges, like all research methods. The strength of PAR as a research methodology stems from various aspects that, within the context of the research we are conducting, produces knowledge that can be applied directly to the local context and to democratise the coproduction of knowledge by collaborating with those most directly affected by the research.[20] The veterans' community, particularly in Wales (the primary geographic region of VLL service users), is characterised by few major metropolitan centres however most of the population live in rural, geographically isolated and lower income areas.[21] As a result, the local knowledge is essential to accurate understanding of problems and the development of effective interventions best suited to provide access to justice for an underrepresented community group and by engaging them through participation in the research process the resulting knowledge is embedded in local contexts.[22] Furthermore, by engaging with the veterans' community in a participatory way we were able to "ensure the relevancy of research questions; increase the capacity of data collection, analysis, and interpretation… and enhance program recruitment, sustainability, and extension".[23] This approach also provided a means for engendering trust[24] and building of community relationships with a community that has a deep mistrust of civilians[25] (discussed in greater detail in section 5.1.2 Ensure ecological

[20] See e.g. M Brydon-Miller, D Greenwood and P Maguire. 'Why Action Research?' 2003 1(1) Action Research, 9; Shortall, 'Participatory action research' in R Miller and J Brewer (eds.), *The AZ of social research* (Sage, 2003), 225.

[21] This is consistent with the view that PAR involves collaborating with individuals from marginalised groups for emancipatory aims see e.g. D Greenwood and M Levin, *Introduction to Action Research: social research for social change* (Sage 2007).

[22] See JN Hughes, 'Commentary: Participatory action research leads to sustainable school and community improvement' (2003) 32(1) School Psychology Review, 39.

[23] J Jagosh, AC Macaulay, P Pluye, JON Salsberg, PL Bush, JIM Henderson, et al., 'Uncovering the benefits of participatory research: implications of a realist review for health research and practice' (2012) 90 (2) The Milbank Quarterly, 311, 312.

[24] C Lenette, N Stavropoulou, C Nunn, ST Kong, T Cook, K Coddington and S Banks, 'Brushed under the carpet: Examining the complexities of participatory research' (2019) 3(2) Research for All, 166.

[25] JD Brewer and S Herron, *Understanding 'Negative Transitioning' in British Ex-Service Personnel* (Queen's University Belfast, 2022).Available at https://www.fim-trust.org/wp-content/uploads/QUB-Negative-Transition-FINAL.pdf

and cultural sensitivity in data gathering).[26] On the other hand, the challenges posed by PAR included participant recruitment particularly when carried out online[27] and the desire for a high level of community involvement.[28] Through our study, we saw first-hand some of the above-mentioned benefits and challenges and discuss them in more detail later.

4. The Research Process

In this paper we reflect on an access to justice platform project which originated from the VLL, a well-established partnership between academics and community partners. At the outset, we should point out that our research project is still ongoing and hence, we present findings from Stages 1 and 2 of our three-stage iterative research process. Data collection and analysis from Stage 1 of the study informed the themes that underpinned Stage 2. Throughout the lifetime of the project, and since its inception, ethics approval for the overall VLL project and for the specific research activities have been granted through Aberystwyth Universities Ethics Board.

In phase 1, data was collected through a combination of co-facilitated focus groups and a web survey. In collaboration with existing community partners we conducted Focus Group Discussions (FGD), we used convenience sampling to generate a heterogeneous sample of veterans (older and younger, men and women, service users and service providers, individuals with varying degrees of technical literacy and geographic regions (North, South, East and West)) and sent participants a copy of the information sheet and a consent form by email. Furthermore, in line with PAR methodology, we chose FGDs over group interviews, as this enabled us as academic researchers to take a peripheral rather than a centre-stage role in the

[26] M Hoffman, 'Between Order and Execution: A Phenomenological Approach to the Role of Relationships in Military Culture' (2020) 6(3) Journal of Veterans Studies, 72.
[27] KK O'Brien, 'Considerations for conducting Web-based survey research with people living with human immunodeficiency virus using a community-based participatory approach' (2014) 16 (3) Journal of medical Internet research e81. 13, doi:10.2196/jmir.3064.
[28] M Viswanathan and others, 'Community-based participatory research: Assessing the evidence: Summary' in *AHRQ Evidence Report Summaries* (Agency for Healthcare Research and Quality (US) 2004). Available from: https://www.ncbi.nlm.nih.gov/books/NBK11852/

focus group discussions.[29] In doing so, our aim was to facilitate group discussions between veteran participants and not to conduct group interviews.[30] Five focus groups were conducted with between 5-10 participants and aimed to collect data surrounding the following themes, veterans use of technology, barriers to accessing online services, perspectives and utility of current services and additional unmet needs for the veteran community. The focus group participants were recruited from the steering groups' service users and were co-facilitated by their organisations (RBL and Change Step) in conjunction with VLL. The co-facilitation was needed due to the high levels of mistrust among veterans towards civilians and were held both digitally (4) and in-person (1). Following Breen's recommendations[31], we devoted the lion's share of our discussion time to probing participants' experiences of using technology and accessing online services, asking them to share and compare their experiences, and discussing the extent to which they agree or disagree with each other. It was not until the final third of the focus groups that we explicitly asked the following questions and facilitated discussion on the topics:

- How do participants use technology currently?;
- What are the barriers to accessing online services?;
- What is the response to a proposed enhanced online access to justice portal and online services?; and
- What other services participants might find useful?

[29] TO Nyumba, K Wilson,CJ Derrick and N Mukherjee, 'The use of focus group discussion methodology: Insights from two decades of application in conservation' (2018) 9(1) Methods in Ecology and evolution, 20.
[30] M Bloor, J Frankland, M Thomas and K Robson, *Focus groups in social research* (Sage 2001).
[31] RL Breen, 'A practical guide to focus-group research' (2006) 30(3) Journal of geography in higher education, 463.

To triangulate and ensure the credibility of the themes from the focus groups two subtypes of closed-ended survey questions, that is Yes/No questions and Likert Scale Multiple Choice Questions (i.e. not important, somewhat important, somewhat useful, very useful) were developed to survey a broader, nationwide sample of veterans.[32] The questions for the survey revolved around the themes that emerged from the focus groups: current provision of access to justice for veterans, preferences in means of accessing legal services and features of an online service that would be most valuable for veterans accessing an online portal to facilitate access to justice. The survey questionnaire was co-designed with the steering group and representatives from VLL partner organisations. Furthermore, we added open-ended questions to the survey questionnaire and developed them in such a way as to give respondents the freedom to give their opinion in their own words thereby adding authenticity, diversity of responses, nuances in opinions and depth and context to the results.[33] These open-ended questions were the following:

- What encourages you to trust a website or mobile app?,
- What other features would you like to see from a free legal advice website or mobile app?; And
- What's your favourite website, and why?.

Our aim in undertaking the survey was to confirm our reflections and observations from the results obtained from the convenience sample of VLL's partnering organisations for the focus groups. The method of recruitment for survey participants included both snowball sampling via partner organisations and social media recruitment of the broader community of

[32] S Roopa and MS Rani, 'Questionnaire designing for a survey' (2012) 46(4_suppl1) Journal of Indian Orthodontic Society, 273.
[33] DA Dillman, JD Smyth and LM Christian, *Internet, phone, mail, and mixed-mode surveys: the tailored design method* (John Wiley & Sons 2014).

veterans in order to reduce selection bias and increase representativeness. Notably, the survey instrument was hosted using JISC Online Surveys (formally Bristol Online Survey). It was decided to distribute the survey via two social media platforms Facebook and LinkedIn. We chose Facebook as it is arguably one of the world's most widely used social media platforms[34] and therefore offers the opportunity to engage and recruit populations traditionally underrepresented in research such as veterans.[35] The procedure was different for each of the platforms. For the distribution of the survey via Facebook we used paid advertising targeted at veterans' communities. The interests that were targeted were individuals with military interests (Royal Air Force, British Armed Forces, Army, Veterans, Marines, Navy, Parachute Regiment (United Kingdom), Support The British Army). The adverts were promoted on side panels and appeared in individuals' news feeds on their home page. The adverts briefly described the study and included a link to the online survey to click through to the online JISC survey. On the other hand, for the distribution of the survey link via LinkedIn, a post was created and posted to the Veterans Legal Link LinkedIn project page that included a brief description of the research and a link to the JISC survey. The approach for the distribution of the survey link was organic without paid advertising. The LinkedIn platform allows for the creation of interest and group pages. Pages and groups to post the survey in were identified by searching the groups using the terms British Army, UK Veterans, RAF, Royal Air Force, and British Marines. We selected eight groups based on membership size (1,000+) to promote the co-designed survey. The post that was created for the VLL's project page was posted on the identified group and interest pages. Additionally, colleagues promoted the post on their own personal LinkedIn feeds. We observed that interactivity (i.e. likes, comments, shares and reactions) helped to build interest in our research and enhanced its credibility and that this ultimately led to an increase in the

[34] S Alhabash and M Ma, 'A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students?' (2017) *Social Media + Society* https://doi.org/10.1177/2056305117691544
[35] ER Pedersen et al., 'Using facebook to recruit young adult veterans: online mental health research' (2015) 4(2) JMIR research protocols e63.

number of participants. The total number of surveys completed through the duration of the survey was 1169, of which 528 were valid responses (discussed in section 5.2.1 Deception by online survey participants).

Furthermore, we used NVivo 12 software for the inductive coding, analysis and interpretation of the open-ended survey question data and followed the following six-step thematic framework recommended by Braun and Clarke[36] to review critically the data and to develop themes:

1. Become familiar with the data
2. Generate initial codes,
3. Search for themes,
4. Review themes,
5. Define and name themes,
6. Produce the report

In addition, intercoder reliability was ensured as two different academic researchers coded the same dataset.[37] After several runs between coders, an intercoder agreement of 100% was eventually achieved for codes and their meanings. Following the analysis of the survey the results were taken back to the steering group in line with Lincoln & Guba's member checking process.[38] The results were consistent with the original focus groups aims and confirmed the reliability and credibility of our findings and interpretations with veteran participants. The

[36] V Braun and V Clarke, 'Using thematic analysis in psychology' (2006) 3(2) Qualitative research in psychology 77.
[37] C O'Connor and H Joffe, 'Intercoder reliability in qualitative research: debates and practical guidelines' (2020) 19 International Journal of Qualitative Methods 1.
[38] YS Lincoln and EG Guba, E. G, *Naturalistic inquiry* (Sage 1985).

robust analysis of data ensured that our study remained credible and that the data could be used to inform and develop the evidence base for the subsequent prototype development.

The insights gathered from Stage 1 informed the action plans for Stage 2 (Developing a prototype of the access to justice platform) of the research that revolves around the development of an online platform to facilitate access to justice for veterans and their families. In order to address and "change [the] social reality"[39] and advance the goals of the VLL a product development map was created[40], inherent within this was the need for a formalised gap analysis and scoping exercise to create a platform development team composed of a web project manager and software designer. Having developed a profound understanding of the lived experiences of end-users of the platform, Stage 2 of our study involved translating these insights into an interactive computer-based prototype. We needed to adapt our research to remote data collection methods due to the Covid-19 pandemic and hence we used remote and digitally enabled ethnographic research methods. This was achieved through the use of "so-called smart devices, e.g., smartphones, tablets and computers, that facilitate[ed] work and enable[d] the understanding of cultural patterns in digital or physical spaces, or in a hybrid intertwining of both digital and physical realm".[41] Specifically, in order to capture qualitative experience data, we used the screen sharing in Microsoft Teams to observe the users experience, interact synchronously with users and interview users in real-time.[42] Remote and digitally enabled ethnographic research methods thus provided us with a realistic user

[39] J Bergold and S Thomas. 'Participatory Research Methods: A Methodological Approach in Motion' (2012) 37(4) Historical Social Research / Historische Sozialforschung 191.
[40] A Liem and EBN Sanders, 'The impact of human-centred design workshops in strategic design projects' in M Kurosu (ed.), *International Conference on Human Centered Design* (Springer, 2011), 110;.AR Lyon, SK Brewer and PA Areán, 'Leveraging human-centered design to implement modern psychological science: Return on an early investment' (2020) 75(8) American Psychologist, 1067. http://dx.doi.org/10.1037/amp0000652
[41] D Podjed, 'Renewal of Ethnography in the Time of the COVID-19 Crisis' (2021) 59(1) Sociology & Space 270.
[42] J Black and M Abrams, 'Remote Usability Testing' in K Norman and J Kirakowski (eds.), *The Wiley Handbook of Human Computer Interaction Set* (John Wiley & Sons 2017) 277; J English and L Rampoldi-Hnilo, 'Remote contextual inquiry: A technique to improve enterprise software' in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Sage 2004) 48 (13) 1483.

environment for conducting prototype testing with users. Participants of this stage of the research were provided detailed information sheets and informed consent with a variety of options to address security and privacy concerns such as the ability to op-in/out of screen sharing, allowing them to opt for a camera off session and to not have the sessions recorded. All of the participants in this stage agreed to screen sharing, camera on, and to have the sessions recorded.

5. Reflections on the application of PAR for the development of an access to justice platform

In the following subsections, we reflect on some of the benefits that accrued from our efforts to apply PAR for the development of an access to justice platform. These benefits are as follows: enhance recruitment capacity, ensure ecological and cultural sensitivity in data gathering and instrumental benefits. On the other hand, we also reflect on the following challenges that we faced as legal academics engaging in a sustained PAR project: deception by online survey participants and involvement of community stakeholders in some but not all phases of the research process.

5.1. Benefits

5.1.1. Enhance recruitment capacity

One of the benefits of PAR is partnership synergy, that is the "combined effect of complementary tangible and intangible partnership assets and enabling processes that gives partnerships unique advantages over the work of individual people or organizations working

towards the same goals".[43] Notably, in our project, recruitment of community members to the steering group was accelerated by the fact that academic-community partners had a long-standing relationship dating back more than 6 years via the VLL project. Thus, existing partners who were representatives of respected veteran community organisations readily accepted the invitation to join the project steering group and hence actively participated in meetings and were strong believers in the project's benefits for the communities they served. Furthermore, from a synergistic perspective, the fact that the project had forged strong links with respected and trusted veterans' community organisations proved critical in the recruitment of additional community representatives to the steering group and the recruitment of veteran participants into our research. Thus, the involvement of community partners enhanced the credibility of the project, increased trust and reduced barriers to recruitment.

5.1.2. Ensure ecological and cultural sensitivity in data gathering

As civilian academic researchers, we were conscious of the fact that for many veterans transitioning from service in the armed forces to civilian life, developing trusting relationships with "civilian outsiders" could be a difficult proposition, as the military is a close-knit community.[44] In our project, through their familiarity with and sensitivity to access to justice issues, veteran community partners who were members of the project steering group were able to ensure ecological and cultural sensitivity in data gathering. Specifically, they contributed invaluable knowledge of the veterans' community by recruiting underrepresented veteran participants and explaining the aims of the project and the involvement of civilian university

[43] E Loban, C Scott, V Lewis, S Law and J Haggerty, 'Activating Partnership Assets to Produce Synergy in Primary Health Care: A Mixed Methods Study' in *Healthcare* (Multidisciplinary Digital Publishing Institute 2021) 9 (8) 1060, 1064.
[44] M Hoffman, 'Between Order and Execution: A Phenomenological Approach to the Role of Relationships in Military Culture' (2020) 6(3) Journal of Veterans Studies 72.

academic researchers to veteran community members. They were also involved in the development of survey questions and co-facilitated all the focus groups. For example, the fact that the total number of valid surveys completed through the duration of our survey was 528 serves as evidence for the cultural acceptability of the survey[45]. Thus, the contribution of veteran community partners was critical to the acceptability and validity of our research.

5.1.3. Instrumental benefits

According to Hagan "[m]ost access-to-justice technologies are designed by lawyers and reflect lawyers' perspectives on what people need. Hence it is not surprising that most of these technologies do not fulfil their promise because the people they are designed to serve do not use them".[46] Our own experience of applying PAR to the generation of ideas for the development of an access to justice platform validates Hagan's statement. It was apparent from our research that veterans were acutely aware of problems with the existing approach to delivering access to justice and as "experts by experience" including them in key phases of the research process helped to facilitate new concept generation for the design and development of the access to justice platform. This collaborative approach has direct implications for the acceptability, uptake and adoption of the platform by veteran end users as the platform is more likely to meet their expectations and requirements. However, as mentioned above, we have completed Stages 1 and 2 of our three-stage iterative research process which comprises the following steps: Understanding and cross-checking the lived experience of the veteran community (Stage 1), developing and testing a prototype of the access to justice platform (Stage 2) and creating the final product and giving real users an opportunity to use the platform

[45] We conducted a qualitative analysis of the survey results however were we to have done a quantitative statistical analysis of the data the sample size would have represented a 99% confidence level with between a 5-6% margin of error.
[46] M Hagan, M 'Participatory design for innovation in access to justice' (2019) 148(1) Daedalus 120.

(Stage 3); and hence it is too early to draw any conclusions about the impact of PAR on the uptake and adoption of the access to justice platform.

5.2. Challenges

5.2.1. Deception by online survey participants

The misrepresentation of survey participants who were eligible to take part in our research was one of the major challenges that we faced within the context of this project. We should first point out that the COVID-19 pandemic forced us to move from what had initially been designed as an in-person survey where the interviewer is physically present to ask the survey questions and to assist the respondent in answering, to an online survey where there is no interviewer present. Although online surveys provide valuable benefits to researchers[47], they are not without limitations as we found out first-hand. Based on our own experience, the benefits of online surveys include the following: ease of capturing open-ended comments, reach and scalability, relatively low cost of administration, speed of distribution, reach and ease of data entry and analysis and the ability to reach sample members who are difficult to reach by other means; and this has been confirmed in the extant literature.[48] On the other hand, one of the main challenges we faced whilst conducting our survey was deceptive practices by participants. These practices were the provision of duplicate responses and misrepresentation of eligibility criteria both of which can be attributed to our decision to offer incentives and the anonymity intrinsic to online surveys; and this is consistent with the existing literature on online surveys.[49]

[47] See e.g. D Andrews, B Nonnecke and J Preece, 'Electronic survey methodology: A case study in reaching hard-to-involve internet users' (2003) 16(2) International Journal of Human Computer Interaction, 185; J McPeake, M Bateson and A O'Neill, 'Electronic surveys: how to maximise success' (2014) 21(3) Nurse Researcher, 24.
[48] HL Ball, 'Conducting Online Surveys' (2019) 35(3) Journal of Human Lactation 413; RE Joel and A Mathur, 'The value of online surveys' (2005) 15(2) Internet Research 195.
[49] HF Lynch, S Joffe, H Thirumurthy, D Xie and EA Largent 'Association between financial incentives and participant deception about study eligibility' (2019) 2(1) JAMA network open e187355; R Pozzar et al., 'Threats of bots and other bad actors to data quality following research participant recruitment through social media: Cross-

Notably, we offered an option to be included in a drawing of five £20 Amazon gift card as an inducement to participate. This incentive was advertised within the social media postings on both our LinkedIn and Facebook platforms. This, we suspect, was the main motivation for non-eligible participants to complete the survey to obtain the financial incentive. This became a larger problem when it became apparent that not only were we receiving individual non-eligible responses but that some respondents were "spamming" (repeatedly filling out the survey) responses. The phenomena of receiving spam responses took several days from the launch of the survey to begin taking place and several additional days for it to be identified that these were potentially mis-represented responses to the questionnaire. Several aspects of the responses constituted a cause for concern. These included the following. First, some of the free text responses that were being received were in languages other than English or with grammatical structures that suggested a non-native English speaker. While this for some surveys may not be indicative of invalid responses, this questionnaire was intended for ex-forces personnel who served in the British armed forces whose language proficiency in English would either be to a native level or a Common European Framework of Reference (CERF) A1 level. Second, surveys were completed in succession over a period of time as indicated by the time stamp for the submission or were completed (from start to finish) in an unbelievably short amount of time, much faster than average reading speeds for the amount of text in the survey. And third, there were duplicated or suspiciously similar responses across the questions as JISC Online Surveys do not prevent a survey from being completed many times on the same computer or from the same IP address unless survey access control is utilised. All of these posed threats to sample validity and data integrity. Therefore, in order to identify and minimize misrepresentation by participants seeking enrolment in our online survey thereby enabling

sectional questionnaire'. (2020) 22(10) Journal of medical Internet research e23021; J Bohannon, 'SCIENTIFIC INTEGRITY: Survey fraud test sparks battle' (2016) 351 Science 6277.

veterans to make their voices heard we used the following combination of strategies recommended in the literature: technical/software strategies and data analytic strategies. With regard to the former, we analysed the dataset to identify unusual or unexpected completion patterns. For example, we found that some respondents were completing the total survey of 18 items and 6 sub-items in 3 minutes or less, whereas the majority of respondents required 6 minutes. In terms of the latter, we determined if data from the sample of respondents who were suspected of misrepresenting their eligibility, differed significantly from the rest of the study sample or if results of the study were substantially different when either including or excluding their data from analyses.[50] Once identified suspected fraudulent results were removed from the sample, leaving 528 survey responses. These responses were the sample that informed the next phase of the PAR cycle.

5.2.2. Involvement of community stakeholders in some but not all phases of the research process

Ideally, PAR calls for the active involvement of community stakeholders as equal partners *in all phases of the research process* from defining relevant research questions, to planning, designing and implementing the investigation, strengthening recruitment strategies, collecting and analysing data and interpreting and applying findings and disseminating outcomes; based on our own experience these requirements are very difficult to meet. This is consistent with Brown's findings as evident in the following statement: "[d]epending on the design participatory research needs to be seen as a continuum from being minimally participatory to being fully egalitarian, whereby realistically most participatory research designs are situated somewhere in between the two with the level of participation changing throughout the

[50] J Kramer et al., 'Strategies to address participant misrepresentation for eligibility in Web-based research' (2014) 23(1) International journal of methods in psychiatric research 120.

process".[51] Thus our PAR design could be described as moderately participatory (having significant participatory elements) as community stakeholders were not involved in defining relevant research questions, collecting and analysing data and interpreting and applying findings. This can be attributed to two interrelated factors. First, the disruption caused by the COVID-19 pandemic led to a shift in priorities on the part of community stakeholders from committing significant time and effort to non-essential activities such as our PAR project to focusing on existential threats and this finding has recently been confirmed by Köpsel, de Moura Kiipper and Peck.[52] In their survey on the impact of the Covid-19 pandemic on stakeholder engagement activities they found that for 45% of respondents, the social distancing measures made it harder to reach stakeholders, and 41% perceived that stakeholders' priorities have shifted away from the research project. Moreover, one-third of participants stated that stakeholders appear to have less time for meetings, be they virtual or physical, than before the start of lockdowns and distancing.[53] However, in this connection it is important to point out that a study by Hayward and colleagues found that: "the choice not to participate can actually be viewed as an act of empowerment [and that]… [a]ssessing social inclusion by measuring levels of participation may therefore be misleading and may not account for community members who have made the rational choice not to participate for any of a number of reasons."[54]And second, following recommendations from Gillis and Jackson[55] we paid sensitivity and attention to veterans' community stakeholders' agenda throughout the research project thereby avoiding any misinterpretation in terms of under- or overestimating their motivation and commitment. Thus, we recognised that their contribution was inextricably

[51] N Brown, 'Scope and continuum of participatory research' (2021) International Journal of Research & Method in Education 2.
[52] V Köpsel, G de Moura Kiipper and MA Peck, 'Stakeholder engagement vs. social distancing—how does the Covid-19 pandemic affect participatory research in EU marine science projects?' (2021) Maritime Studies 1.
[53] ibid 9.
[54] See e.g. C Hayward, L Simpson and L Wood 'Still left out in the cold: Problematising Participatory Research and development' (2004) 44(1) Sociologia Ruralis, 100.
[55] A Gillis and W. Jackson, *Research for nurses: Methods and interpretation* (FA Davis Company 2002).

linked to what Hayward and colleagues refer to as "self-defined boundaries"[56] and that these boundaries could fluctuate depending on circumstances and contexts and hence it was necessary for us to be flexible and ready to adapt to possible changes during the research process.

6. Conclusion

In this paper, we reflected on our experiences of developing the UK's first access to justice platform for veterans and their families through an ongoing PAR project that brought together armed forces veterans, representatives from veterans' service providers, and the Veterans Legal Link team members comprising of legal academics, lawyers, sociologists, computer software designers and graphic designers to collect, interpret, and apply community information to address issues related to the delivery of access to justice. We presented findings from Stages 1 and 2 of our three-stage iterative research process which included the following steps: Understanding and cross-checking the lived experience of the veteran community (Stage 1), developing and testing a prototype of the access to justice platform (Stage 2) and creating the final product and giving real users an opportunity to use the platform (Stage 3). Data collection and analysis from Stage 1 of the study informed the themes that underpinned Stage 2. As demonstrated by the foregoing discussion, data was collected through the following methods: co-facilitated focus group discussions, a web survey that was codesigned with veteran community stakeholders and remote and digitally enabled ethnographic research methods. We included several reflections that may help legal practitioners and researchers interested in applying PAR within the area of access to justice and the field of legal research. As discussed above, some of the benefits that accrued from our efforts to apply PAR for the development of

[56] See Hayward et al (n 54).

an access to justice platform included the following: enhance recruitment capacity, ensure ecological and cultural sensitivity in data gathering and instrumental benefits. On the other hand, we also faced the following challenges: deception by online survey participants and involvement of community stakeholders in some but not all phases of the research process. The foregoing begs the following question: Is PAR an approach to research for law? The answer, of course, depends upon the nature of the research. If the proposed research focuses on bridging the gap between theory and practice and is aimed at solving concrete social problems, then PAR with its capacity to surface social, political and cultural issues and generate local knowledge that can inform practical solutions and actions for social transformation, would be a wise choice.

Journal of Legal Research Methodology

Enhancing the Measurement of Sentence Severity through Expert Knowledge Elicitation

Jose Pina-Sánchez · John Paul Gosling

Abstract Quantitative research on judicial decision-making faces the methodological challenge of analysing disposal types that are measured in different units (e.g. money for fines, days for custodial sentences). To overcome this problem a wide range of scales of sentence severity have been suggested in the literature. One particular group of severity scales that has achieved high validity and reliability are those based on Thurstone's pairwise comparisons. However, this method invokes a series of simplifying assumptions, one of them being that the range of severity covered by different disposal types is constant. We undertook an expert elicitation workshop to assess the validity of that assumption. Responses from the six criminal law practitioners and researchers that participated in our workshop unanimously pointed at severity ranges being highly variable across disposal types (e.g. much wider severity ranges were identified for suspended custodial sentences than for fines). We used this information to re-specify Thurstone's model allowing for unequal variances. As a result, we obtained a new, more robust, scale of sentence severity.

Keywords Sentencing · severity · Thurstone scale · expert elicitation

This work was supported by the National Centre for Research Methods [RIS14241/06]. We are also immensely grateful to the six sentencing experts that participated in our workshop: Julian Berg, Elizabeth Bourgeois, David Hayes, Eleanor Nicholls, Ruth Pope, and Sebastian Walker. This study would not have been possible without the time and unique subject knowledge that they so generously offered.

J. Pina-S´anchez
School of Law, University of Leeds Liberty Building, University Western Campus, Moorland Rd, Leeds LS3 1DB, UK
ORCID: 0000-0002-9416-6022
E-mail: j.pinasanchez@leeds.ac.uk

J.P. Gosling
School of Mathematics, Durham University, UK
ORCID: 0000-0002-4072-3022

1 Introduction

Most quantitative sentencing research seeks to explore the causal mechanisms involved in judicial decision-making, a task that is normally undertaken by studying the variability in the relative severity of large samples of sentences. Compared to other legal or criminal justice research areas, the formal setting in which the sentencing process takes place simplifies its analytical complexity. For example, the outcome variable (i.e. the sentence imposed) is often unequivocally recorded (i.e. measured without error), reverse causality is non-existent since the path from case processing to sentence is unidirectional, and although problems of unobserved confounders are as common as in any other area relying on observational data, sentencing researchers rarely fail to acknowledge this problem and often anticipate its biasing effect. There is, however, one particular methodological challenge affecting sentencing research specifically. This problem stems from the fact that judges can rely on a wide range of disposal types (i.e. sentence types, or sentence outcomes), which are not readily comparable, since they are measured in different units (Freiberg and Fox, 1986). For example, fines can be expressed in pounds, while custodial sentences are measured in days, while community orders are normally based on the completion of certain conditions.

Researchers have often addressed this problem by either restricting their analysis to one disposal type - commonly custodial sentence length - or by specifying the probability of custody compared to other possible outcomes. The former approach induces selection bias (Bushway et al., 2007) since custodial sentences are the most severe of all possible disposal types, representing a small part of all sentences imposed (roughly 7% of the total in England and Wales). The latter involves a substantial loss of information, since it takes all non-custodial outcomes as a homogeneous group, rendering sentencing analyses unduly blunt and leading to forms of measurement error (Berkson, 1950). More methodologically advanced researchers have sought to adjust fore selection bias using Tobit (Albonetti, 1998; King et al., 2010; Kurlychek and Johnson, 2010), hurdle (Hester and Hartman, 2017), or Heckman's two-stage (Feldmeyer and Ulmer, 2011; Steffensmeier and DeMuth, 2001; Ulmer et al., 2010) models. However, these models are also based on different assumptions, e.g. that the type and quantity of the sentence outcome are decided in different steps of the sentencing process, or that the unobserved severity of non-custodial outcomes stems from a hypothetical distribution, which observed, right-hand side is represented by the length of custodial sentences. Pina-S´anchez and Gosling (2020) demonstrated how, at least for the case of England and Wales, those two assumptions are violated. Perhaps more importantly, all of the statistical adjustments that have been suggested in the literature, involve discarding any variability recorded across non-custodial outcomes.

An alternative strategy relies on adopting a scale of sentence severity (Leclerc and Tremblay, 2016; Yan and Lao, 2021). This involves assuming an underlying continuum along which sentence outcomes can be located; so they can all be expressed under the same measurement unit, preventing any loss of information. Unlike the more formally defined statistical adjustments used in the literature, the estimation of scales of severity represents a widely heterogeneous approach, manifested in multiple types of solutions, none of them without limitations. The most common approach is magnitude escalation, which involves establishing a reference sentence, and deriving the relative severity of other sentence outcomes from the subjective comparisons made by a sample of informed participants (Harlow et al., 1995; Spelman, 1995; Tremblay, 1988). In our view, the main problem affecting these types of studies stems from the wide variability in participants' responses, making scales highly sensitive to the composition of the sample studied, which compromises their reliability. Other studies have relied on data-driven methods such as correspondence analysis (Francis et al., 2005; McDavid and Stipack, 1981). Although these methods remove the inherent unreliability of subjective perceptions, they derive the relative severity of different sentence outcomes from the frequency with which they are used as punishments for different crime types. This in turn invokes further assumptions, such as perfect proportionality between crime seriousness and sentence severity, which, if violated can lead to nonsen-

sical severity scores, such as longer suspended sentences deemed more severe than shorter ones; hence, questioning the validity of such data-driven scales.

Our preferred approach for the estimation of sentence severity is Thurstone (1927) pairwise comparisons. This method is based on subjective perceptions (Buchner, 1979; Pina-S´anchez et al., 2019a), but rather than requesting research subjects to report the magnitude by which a sentence outcome is more severe than another, it relies on either ordinal comparisons, or alternatively, on relative comparisons requesting how often - rather than how much - one sentence outcome is more severe than another. That is, the cognitive burden placed on research participants is eased by asking them to identify which of the two outcomes is more severe, or by framing the comparison in terms of frequencies.[1] However, the method is also reliant on a series of parametric assumptions. The potential severity associated with each sentence outcome is assumed to be normally distributed. The means of these distributions vary, reflecting the severity scores attributed to each sentence outcome, but their variances are assumed to be equal. This is a convenient assumption that makes the estimation process more parsimonious. We should nonetheless question its validity, since it involves assuming that the range of severity scores covered by each sentence outcome is equivalent across all of them. For example, using the standard Thurstone model, we might find that community orders are in general more severe than fines, but we will have to assume that the difference between the minimum and maximum severity attributed to each of these outcomes is the same. Importantly, if this assumption is violated, the severity scores estimated for each sentence outcome will be biased, as they won't be accurately reflecting their relative distance in the underlying scale of severity.

In this article we demonstrate how the assumption of equal variances can be relaxed using expert knowledge elicitation techniques. We do so by revisiting the Thurstone scales of severity estimated in Pina-Sanchez et al. (2019a) and Pina-S´anchez and Gosling (2020), using qualitative insights elicited from six criminal law experts and an extended version of the standard Thurstone model. As such, our article contributes to both the measurement and sentencing literature in three important ways: i) we show the importance of testing the underlying assumptions of Thurstone scales; ii) demonstrate the multiple advantages of relying on expert elicitation methods for the estimation of sentence severity; and iii) estimate a more robust scale of severity with which to undertake sentencing research in the jurisdiction of England and Wales. The creation of this new scale of severity represents a timely contribution as studies employing severity scales are becoming more widespread (Roberts and Bild, 2021; Isaac, 2021), but also as new and more sophisticated sentencing datasets that capture differences across non-custodial sentences in unprecedented detail, are becoming increasingly available (Sentencing Council, 2021; Ministry of Justice, 2021).

2 Challenging the Assumption of Equal Variances

The Thurstone model works by linking each sentence outcome to a corresponding set of latent normal distributions that have the following property: the probability of a random draw from distribution A exceeding a random draw from distribution B is equal to the proportion of times sentence outcome A will be preferable to sentence outcome B (Thurstone, 1927; Pina-S´anchez and Gosling, 2020). The simplest - and most commonly employed - adaptation of the Thurstone model is in its 'Case V' form (Mosteller, 1951). Under this setting, the latent normal distributions for each sentence outcome are assumed to have equal variance of a half, so that the differences between sentence outcomes have a variance of one. Given a set of judged preference proportions, the Thurstone model is fitted and the means of the resulting latent normal distributions can be used as a ranking or scoring metric. Lastly,

[1] By requesting frequency rather than magnitude comparisons, responses are framed within a 0 to 1 range - as opposed to the 0 to ∞ range involved in magnitude comparisons - which provides intuitive points of reference at 0 (never more severe), 0.5 (as severe) and 1 (always more severe).

because the choice of variance of the latent variables is arbitrary, it is valid - and common - for the metrics to be rescaled to fit a more convenient range (0-100, for instance).

This assumption of equal variances is possibly robust enough in many contexts where Thurstone scaling is employed, but should be questioned when we consider the relative severity of different sentence outcomes. We argue that the range of severity that could be covered by different disposal types is likely proportional to their heterogeneity. For example, fines can only be expressed in pounds, and have specific bands associated to them, whereas community orders or suspended sentences could be composed of a wide range of conditions, such as curfews, completion of rehabilitative programs, unpaid work, and they could also include fines too. Therefore, given that fines are just a subset of the possible sentence outcomes that could make part of a community order sentence, it is not tenable to see the two disposal types as covering equal ranges of potential severity, i.e. the former will always be narrower than the latter. Once established the lack of theoretical soundness of this assumption, it is worth considering how exactly could results from the Thurstone scale (severity scores for different sentence outcomes) be affected.

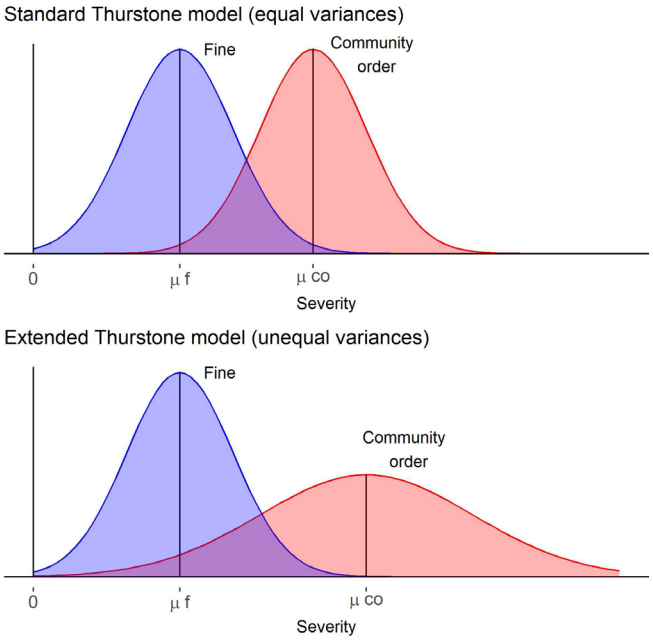


Fig. 1 Simplified representation of the Thurstone scaling method under the assumption of equal variances (top) and allowing for different variances (bottom).

The implications of violating the hypothesis of equal variances is shown visually in Figure 1, based on a simple - made up - example showing the severity scores for two disposal types, fines and community orders, where pairwise comparisons found the latter leading to more severe outcomes than the former in 80% of instances where any of these two disposal are imposed. As previously noted, the severity score for each disposal type (μ_F for fines and μ_{CO} for community orders) is determined by the mean of their respective latent distributions, with those distributions situated as far apart from each other as indicated by the estimates of relative severity derived from their pairwise comparisons. For example, if community orders are considered more severe than fines 50% of times, the two distributions, and their respective means, would be placed at the exact same location (i.e. complete overlap), while a result of

100% would place the minimum severity in the distribution of community orders to follow exactly the maximum severity in the distribution of severity of fines (i.e. no overlap).

Notice as well how it is not just the extent of the overlap between outcomes that determines their mean score, but also the shape of the distributions that they are assumed to follow. The top part of Figure 1 shows the result for the standard Thurstone model, where all outcomes are assumed to be normally distributed with equal variances. The bottom part shows how, for the same level of overlap as on the top graph (20%), the severity score for community order is now higher than before as a result of having considered a latent distribution with a standard deviation twice larger than that used to describe fines. In sum, the severity scores derived from Thurstone scales should be seen - at least in principle - as highly sensitive to the underlying ranges of severity assumed for each of the sentence outcomes considered. We now proceed to investigate such proposal empirically.

3 The Expert Elicitation Workshop

To explore the extent to which ranges of severity vary across different sentence outcomes, we organised an expert elicitation workshop. Simply put, expert elicitation techniques are similar to focus groups with two main distinctions: i) the participants are experts on a given field; and ii) the goal is to retrieve estimates for one or a series of well-defined numerical parameters, normally taking the form of the probability of a given event, or a probability distribution across multiple outcomes (O'Hagan et al., 2006). In addition, to aid repeatability and to help avoid cognitive and social biases, structured elicitation sessions are devised around well-design protocols.

In our case, the Sheffield elicitation protocol was employed, as it is specifically designed to facilitate group judgements about complex quantities whilst recording key reasoning and evidence used by the experts in forming their judgements (Gosling, 2018). The Sheffield elicitation protocol has been used extensively in many areas of science (Dessai et al., 2018; Jansen et al., 2020; Booth and Thomas, 2021) and policy making (Gosling et al., 2012; Usher and Strachan, 2013; Brennan et al., 2017). The outcome of our structured elicitation exercise was a set of judgements on sentencing severity along with documented reasoning, giving ownership to the entire group, i.e. the group discussions and subsequent recording were conducted under the Chatham House rule.

The workshop took place on the 4th of December 2018, at the London Mathematical Society (Morgan House, Russell Sq). All six experts initially approached accepted our invitation. The group of experts was designed to meet two criteria: i) each individual participant should hold expert knowledge about the sentencing process in England and Wales; and ii) the group should reflect different forms of applied and theoretical expertise, including policy-makers, academics, and practitioners. In alphabetical order, the participating experts in our workshop were: Julian Berg (Criminal Law Solicitor's Association), Elizabeth Bourgeois (Bradford and Keighley Magistrates Court), David Hayes (University of Sheffield), Eleanor Nicholls and Ruth Pope (both from the Sentencing Council for England and Wales), and Sebastian Walker (Law Commission). The authors of this article conducted the workshop, which was divided in two parts, for a total of four hours.

The first and longest part of the workshop involved eliciting estimates of the severity overlap across different pairs of sentences. To frame the group discussions we posed specific questions. However, given the complexity of the topic, we formulated the main question in two different ways. We asked: i) *How often can sentence-X be more punitive than sentence-Y?* ; and ii) *What proportion of offenders would prefer sentence-X than sentence-Y?* To provide further context to the discussion, we also asked our participants to consider the following: i) the heterogeneity of sentences possible within a given disposal type (e.g. the different conditions that could be attached to a long community order); ii) that we are not asking whether the different sentences can be used interchangeably, but rather whether there are circumstances when one can have a more punitive effect than the other; and iii) to consider not just the average offender, but the mix of different offenders seen through courts. Discussions for each of the

pairs of sentence outcomes compared were allowed to take as long as necessary, until a consensus was reached and a specific estimate could be derived.

The format used represents a substantial departure from the typical approach followed in applications of Thurstone's scaling. In its original form, overlaps between categories are estimated by repeating the same pairwise comparison over a large sample of participants using questionnaires. Here, we are asking our participants to identify not only the most severe of two outcomes, but we also request them to estimate how often they think that would be the case. As a result, the questions asked are more cognitive demanding, which is why they are ideally explored under a format that allows for in depth discussion. There is therefore a trade-off between external and internal validity. Such format cannot be easily scaled up to obtain a large sample of expert views. However, given the complexity of the questions asked, we follow Bolger (2018) and O'Hagan (2019) in setting a low number of experts, since the possibly lower reliability associated with a small sample is more than offset from the gains in focus and accuracy made possible through an expert elicitation workshop.

The number of pairs of sentence outcomes compared expanded those considered in Pina-S´anchez et al. (2019a). This was to capture the higher granularity with which non-custodial sentences are recorded by the latest data releases from the Sentencing Council for England and Wales, which disaggregates fines within six bands (A, B, C, D, E, and F), reflecting different quantities, and community orders in three categories (low, medium and high), based on the conditions that could be imposed.[2] This gave us eighteen outcomes to be compared, meaning 153 potential pairwise combinations. To limit participants' fatigue, we restricted our discussion to comparisons of 30 pairs of sentences where it could not be ruled out that one of the sentences would always be deemed more severe than the other, i.e. a severity overlap could be theoretically possible.

In the second part of the workshop we explored the extent to which severity ranges covered by each of the eighteen sentence outcomes explored could be considered equivalent across all of them, and if not, how different could they be. This is an even more complex question, hence, we approached it in two stages. First, we asked introductory questions requesting to identify the sentence outcome with a wider severity range out of a series of pairwise comparisons. After corroborating our working hypothesis (i.e. severity ranges vary widely across different sentences), we proceeded to estimate the relative range for each of the sentences considered. To do so we asked two questions: i) *Do you think [introduce specific disposal type] has a different 'severity spread' than the average disposal type in this list (e.g. than a medium community order)?* ; and ii) *Roughly, how much do you think the spread of severity differs from the average?*; to which the following list of answers were presented [3]: i) *A quarter*

[2] The quantity associated for each of the six fine bands can be found here: https://www.sentencingcouncil.org.uk/explanatory-material/magistrates-court/item/fines-and-financial-orders/approach-to-the-assessment-of-fines-2/2-fine-bands/; the types of conditions that could be attached to different forms of community orders are explained here: https://www.sentencingcouncil.org.uk/droppable/item/community-orders-table/. For a more detailed description of the different disposal types available to sentencers in England and Wales see Harris and Walker (2021).

[3] By comparing relative spread across sentence outcomes we are implicitly assuming that the underlying severity scores for each outcome are part of a ratio-level - rather than interval-level - variable. This implies considering our severity scale as bounded on the left by zero (representing 'no punishment'), but also that all levels of severity along the scale can be objectively defined. The latter is clearly not a realistic assumption since levels of severity are subjectively defined, and as such, direct comparisons of their intensity (e.g. outcome A is twice as severe as outcome B) are bound to be unstable across subjects. See for example Thomas et al. (2018), who demonstrated how, in similar workshops seeking to elicit arrest risks across crime scenarios, participants provided stable probabilistic perceptions of risk that were rank-stable within participants, but were also simultaneously arbitrary in the sense that the specific risk of arrest for each scenario was neither stable between individuals nor meaningful, a set of measurement properties that they deemed to reflect *coherent arbitrariness*. We argue, however, that to assume our severity scale possesses the characteristics of a ratio-level variable is well justified, as the reason we do so is to be able to relax an even less tenable assumption, namely that of equal variances. Put differently, while the assumption of seeing a severity index as a ratio-level variable is theoretically questionable, and the elicitation of comparisons of severity spread methodologically challenging, the assumption of equal variances is not just wrong but directly leads to biased estimates of severity derived from the Thurstone method.

of the average spread (25%); ii) *Half of the average spread (50%)*; iii) *About equal spread to the average (100%)*; iv) *Fifty percent wider than the average spread (150%)*; v) *Twice the average spread (200%)*; vi) *Three times the average spread (300%)*; vii) *Other*.

4 Results

Results from the pairwise severity comparisons are summarised in Table 1. Each cell reports the proportion of instances in which the severity of the sentence outcome on the top of the column could be deemed more severe than the corresponding sentence outcome on the left margin. As could be expected, most sentence outcomes at the two extremes of the distribution (such as discharges, or immediate custodial sentences) show little to no overlap with others, whereas the picture is muddier when we focus on sentences lying on the middle of the distribution, such as fines, community orders and suspended sentences. Notice for example how a band-F fine was identified to be, in some instances, more punitive than a suspended sentence, or how high community orders could similarly be judged, sometimes, more punitive than an immediate custodial sentence. These overlaps resonate well with the concept of 'penal exchangeability' (Freiberg and Fox, 1986; Lovegrove, 2001; Sebba and Nathan, 1984), which sees different disposal types not as discrete steps but as a range of choices that are not always differentiated by their punitive effect, but potentially by other sentencing goals such as restitution, rehabilitation, or incapacitation. Similarly, such overlaps in severity allow acknowledging the concept of 'penal subjectivity', that is, the variability in between-subject experiences of punitiveness resulting from identical sentences (Ginneken and Hayes, 2017; Hayes, 2016, 2018; Padfield, 2011).

Estimates of the variability of severity ranges across sentence outcomes are shown in Table 2. Compared to the reference case (medium community orders), our participants manifested that fines and immediate custodial sentences comprise a much narrower severity range, from around a third to a fourth of the range encompassed by medium community orders. Contrary to that, high community orders and suspended sentences where deemed to cover wider severity ranges, reflecting the ample discretion that judges could apply to configure such disposal types, and the much wider range of punitive outcomes that could arise as a result, e.g. from relatively lenient suspended sentences that will never be activated, to those composed of multiple punitive conditions that will potentially become activated, involving prison time. These results corroborate our suspicion regarding the implausibility of the equal variances assumption invoked in the standard Thurstone model when applied to the estimation of sentence severity.

To assess the implications of violating the equal variances assumption we compare two scales of severity; both of them derived from the pairwise severity comparisons shown in Table 1. However, one scale relies on the standard Thurstone model, assuming equal variances in the severity distributions of each sentence outcome considered, while the other incorporates the different expert elicited variances reported in Table 2.[4] The two scales are reported in Table 3. In both of them, severity scores for immediate custody sentences longer than three months were estimated as a second step by extrapolating linearly from the severity scores obtained from the Thurstone model for one, two, and three months custodial sentences.[5]

The most noticeable effect appears to be the wider range of severity now covered by community orders and suspended sentences, reflecting the substantial heterogeneity that characterises many of

[4] To allow for unequal variances the Thurstone model was estimated from scratch using R, as opposed to relying on built-in functions (e.g. *thurstone*, available in the package *psych* (Revelle, 2018)). The code used to account for unequal variances has been included in the Technical Appendix: R Code at the end of this article.

[5] Non-linear functions were also considered to reflect the marginally diminishing returns of severity that could be expected for every additional month in prison (Leclerc and Tremblay, 2016; Spelman, 1995). To do so, different rates of decay were considered based on insights elicited from our experts. We found that severity scores for sentences longer than five years varied widely depending on the rate of decay considered. Hence, to avoid introducing such a potential source of unreliability, we decided to employ somehow less realistic, but possibly more robust, linear extrapolations.

Table 1 Pairwise severity comparisons (each cell reports how often the sentence at the top of the column could be deemed more severe than the corresponding sentence on the left margin).

	absolute discharge	cond. discharge	fine A	fine B	fine C	fine D	fine E	fine F	low community order	medium community order	high community order	1m custody 6m suspended	1m custody 12m suspended	6m custody 6m suspended	12m custody 24m suspended	1m immediate custody	2m immediate custody	3m immediate custody
absolute discharge	0.50	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
conditional discharge	0	0.50	0.65	0.80	1	1	1	1	1	1	1	1	1	1	1	1	1	1
fine A	0	0.35	0.50	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
fine B	0	0.20	0	0.50	1	1	1	1	1	1	1	1	1	1	1	1	1	1
fine C	0	0	0	0	0.50	1	1	1	0.65	1	1	1	1	1	1	1	1	1
fine D	0	0	0	0	0	0.50	1	1	0.55	0.75	1	0.75	1	1	1	1	1	1
fine E	0	0	0	0	0	0	0.50	1	0.40	0.60	0.95	0.70	0.75	1	1	1	1	1
fine F	0	0	0	0	0	0	0	0.50	0.20	0.50	0.90	0.60	0.65	0.80	1	1	1	1
low community order	0	0	0	0	0	0.45	0.60	0.80	0.50	1	1	1	1	1	1	1	1	1
medium community order	0	0	0	0	0	0.25	0.40	0.50	0	0.50	1	0.55	1	1	1	1	1	1
high community order	0	0	0	0	0	0	0.05	0.10	0	0	0.50	0.40	0.45	0.60	1	0.60	0.70	0.75
1m custody 6m suspended	0	0	0	0	0	0.25	0.30	0.40	0	0.45	0.60	0.50	1	1	1	1	1	1
1m custody 12m suspended	0	0	0	0	0	0	0.25	0.35	0	0	0.55	0	0.50	0.95	1	1	1	1
6m custody 6m suspended	0	0	0	0	0	0	0	0.20	0	0	0.40	0	0.05	0.50	1	0.85	1	1
12m custody 24m suspended	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.50	0.30	0.35	0.40
1m immediate custody	0	0	0	0	0	0	0	0	0	0	0.40	0	0	0.15	0.70	0.50	1	1
2m immediate custody	0	0	0	0	0	0	0	0	0	0	0.30	0	0	0	0.65	0	0.50	1
3m immediate custody	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0.60	0	0	0.50

Table 2 Severity ranges covered by each sentence outcome relative to a medium community order.

sentence outcome	relative range
conditional discharge	62.5%
fine A	37.5%
fine B	37.5%
fine C	37.5%
fine D	37.5%
fine E	37.5%
fine F	37.5%
low community order	100%
medium community order	100%
high community order	125%
1 month custody 6 months suspended	100%
1 month custody 12 months suspended	125%
6 months custody 6 months suspended	125%
12 months custody 24 months suspended	175%
1 month custody	25%
2 months custody	25%
3 months custody	25%

Table 3 Severity scales considering equal and unequal variances.

sentence outcome	equal variances	unequal variances	unequal variances rescaled
absolute discharge	0	0	0
conditional discharge	2.75	1.67	1.47
fine A	3.02	1.83	1.59
fine B	4.32	2.35	2.03
fine C	7.08	4.10	3.55
fine D	8.44	4.79	4.15
fine E	9.26	10.42	9.01
fine F	9.98	10.88	9.42
low community order	8.59	4.86	4.21
medium community order	9.66	10.65	9.22
high community order	11.06	12.26	10.61
1 month custody 6 months suspended	9.90	10.87	9.40
1 month custody 12 months suspended	10.69	11.65	10.08
6 months custody 6 months suspended	11.40	12.37	10.70
12 months custody 24 months suspended	16.56	14.09	12.2
1 month custody	12.01	12.97	11.22
2 months custody	15.65	13.41	11.60
3 months custody	16.59	13.82	11.96
12 months custody	37.65	17.69	15.31
5 years custody	147.58	38.30	33.14
10 years custody	284.98	64.05	55.43
20 years custody	559.80	115.55	100

them, which under the standard form of the Thurstone model appears to be oversimplified. Similarly, we can also observe how the scale accounting for unequal variances show a much narrower range of severity scores across the less punitive fines (fines A, B, and C, those which were never deemed to overlap with more punitive disposal types like community orders), and also across custodial sentences. Specifically, severity scores for fines A (1.83) to C (4.10) are limited to a range of 2.27 when considering unequal variances, but that range expands to 4.06 if equal variances are assumed. Comparisons for the range of severity scores for the first three months of custodial sentences are even starker, 0.85 when allowing for unequal variances, compared to 4.58 when equal variances are assumed. These

relative changes in severity scores across different sentence outcomes reflect well the insights about the variability in severity ranges elicited from our sample of experts.

The much narrower ranges of severity estimated for the three immediate custodial sentences (one to three months) is particularly relevant, since all other custodial sentences are extrapolated from the trend seen in the first three months. As such, being able to estimate the marginal increase in severity associated with an additional month in custody is key in order to obtain valid estimates of severity for all other custodial sentences This effect can be noticed by contrasting the divergence in severity scores between the scales based on equal and unequal variances as custodial sentences grow longer. To facilitate interpretations of the relative severity for different sentence outcomes, Table 3 also includes a rescaled version (ranging from 0 to 100) of the unequal variances scale of severity. For example, using that rescaled index, we can see that the highest possible fines (fine-F), medium community orders, and the shortest possible suspended sentences (1 month custody, 6 months suspended), are all three considered of similar severity, which is also roughly equivalent to twice the severity attributed to low community orders.

5 Discussion

The new scale of severity presented here expands those recently developed by Pina-S´anchez et al. (2019a) and Pina-Sanchez and Gosling (2020) in three important ways.[6] First, we have contemplated a wider range of non-custodial sentences, including different types of fines and community orders. This new capacity to discriminate between non-custodial sentences in higher detail matters, since they represent roughly 93% of sentences imposed. Yet, they rarely feature in quantitative sentencing research, leading to a deeply partial understanding in relation to many of the key questions explored in the discipline, such as: i) what was the impact of key sentencing reforms (Fleetwood et al., 2015; Pina-S´anchez and Linacre, 2014; Roberts and Pina-S´anchez, 2021; Ulmer et al., 2011); ii) the effect of specific aggravating or mitigating factors (Irwin-Rogers and Perry, 2015; Lightowlers and Pina-S´anchez, 2017; Lightowlers et al., 2020; Kane and Minson, 2022); iii) changes in trends of sentence severity (Allen, 2016; Doob and Webster, 2003; Hindelang et al., 1975; Pina-S´anchez et al., 2016); iv) consistency in sentencing (Barbora et al., 2012; Isaac, 2020; Brunton-Smith et al., 2020; Dr´apal, 2020); v) or disparities associated with different offender's demographic characteristics (Baumer, 2013; Yan and Lao, 2021; Isaac, 2020; Pina-Sanchez et al., 2019b); to name a few.

Second, we have relaxed the assumption of equal variances invoked in the standard Thurstone model. That is, we have acknowledged that the ranges of severity which could be attributed to different sentence types are not uniform. This provides a sounder theoretical foundation to the new index of severity presented here, since we can now reflect not only the highly variable conditions that could be attached to different disposal types (e.g. community orders being much more heterogeneous in their potential composition than fines, which are limited to prescribed financial amounts), but we can also reflect the variability in the punitive effect that a given sentence can have across different subjects. Importantly, by improving the theoretical foundations of the scale of severity in such way, we have further enhanced its capacity to accurately discriminate across non-custodial sentences.

Third, to derive the expected differences in severity ranges across sentence outcomes we have employed expert elicitation techniques, a new and promising research design in the penal metric literature. Specifically, we conducted a four hours workshop with six sentencing experts following the Sheffield elicitation protocol (Gosling, 2018). The small sample size required to conduct an expert elicitation workshop of these characteristics effectively has undoubtedly affected the reliability of our findings, however, we believe this is a price worth paying given the higher validity afforded by this research

[6] The latest of which has also been adopted by the Sentencing Council for England and Wales in impact assessments of their sentencing guidelines (Isaac, 2021).

design. In contrast with pairwise comparisons of sentence severity elicited from questionnaire (Pina-S´anchez et al., 2019a; Buchner, 1979; Spelman, 1995), expert elicitation techniques can provide the necessary context, discussion, focus, and time for reflection. As a result, responses are substantially more thoughtful, which, for a subject as complex and subjective as that of sentence severity, represents a much desirable trait.

We also believe that our choice of sentencing experts, which combined academics, members of the bar, sentencers, and policy-makers, is particularly fitting to the types of research questions contemplated during the elicitation workshop. In particular, the group of experts demonstrated ample knowledge regarding the typical conditions that are commonly used for each of the sentence outcomes contemplated. In the literature on penal metric theory, it is more common to see perceptions of sentence severity derived from samples of offenders (McClelland and Alpert, 1985; Petersilia and Deschesnes, 1994; Spelman, 1995) or members of the general public (Erickson and Gibbs, 1979; Tremblay, 1988). We believe, however, that the general population is not sufficiently informed to provide valid answers when queried about sentencing. For example, 17% of respondents in a recent opinion poll conducted for the Scottish Sentencing Council (Black et al., 2019) claimed that the adequate response for an early guilty plea should be an increase in severity as it represents an admission of guilt, while the latest opinion poll conducted by the Sentencing Academy (Roberts et al., 2022) reported that 56% of participants thought the average custodial sentence length in England and Wales was shorter in 2021 than two decades before, even though the average sentence length had actually grown by roughly 50% during that period. We could also expect offenders' perspectives to be biased. Offenders found guilty will know better than anyone else the true severity of the sentence that was imposed on them, but their understanding will likely be limited to that specific sentence - or range of sentences in the case of recidivist offenders. Hence, we would not expect them to be more knowledgeable than other members of the general public when it comes to assess the severity that could be attributed to other sentence types. Nor will they be better equipped to assess the severity which that same sentence could exert on others.

There is, however, one group of experts that we believe are particularly well equipped to provide highly meaningful views in relation to the specific concept of penal subjectivism (i.e. how personal context can make a given sentence more or less severe). These are criminal solicitors and barristers. Their unique understanding of the question at hand stems from their experience in discussing potential sentence outcomes with their clients, which will often involve an explicit revelation of preferences. For example, during discussions on whether to plead guilty. Furthermore, criminal lawyers, especially those that are more senior, will likely have become aware of a higher number of cases, contexts and offenders, providing them a unique perspective in relation to the varying punitive effect of different disposal types. As such, we believe it would be interesting to assess whether the results obtained here would replicate if the expert elicitation workshop was to be conducted with a sample entirely composed of criminal lawyers. Future replications would also provide useful insights into the generalisability of the responses elicited from our sample, and in so doing assess the reliability of the severity scale derived from those responses. Undertaking such work will be especially meaningful following processes of sentencing reforms, or similar structural changes within the criminal justice system, such as the renationalisation of probation services, or any further deterioration of prison conditions. All of them instances where the relative severity of disposal types available to sentencers should be reassessed.

Considering avenues of research with which to continue enhancing the robustness of future severity scales, we identify two areas that ought to receive further attention. First, because longer custodial sentences will always be more severe than shorter ones, there is a limit on the range of severity scores that can be directly estimated using Thurstone pairwise comparisons. In our view, after the assumption of equal variances, this represents the next major limitation affecting the estimation of severity scores using Thurstone scaling. The solution adopted here, based on the linear extrapolation of severity scores of one, two and three months custodial sentences, makes the estimation of severity scores for custodial sentences of four months or longer, more unreliable than for the rest of outcomes directly

estimated through Thurstone pairwise comparisons. Moreover, this uncertainty is not uniform across custodial sentences, but proportional to their length, making long sentences particularly unreliable. This limitation could be potentially overcome by combining more than one estimation method for different parts of the severity scale, e.g. Thurstone scaling up to three months in custody combined with magnitude escalation beyond that.

Lastly, we also intend to explore the use of interactive visual apps. We believe expert elicitation techniques to be the most appropriate approach for the exploration of sentence severity, given the discursive yet highly focused setting that they can facilitate, both features equally necessary to structure discussions around a subject as complex as this. However, the ultimate purpose of expert elicitation is the translation of subjective perceptions into specific numerical parameters, which is something that legal experts are not adequately trained to do. Visual aids could be employed to facilitate that process. In particular, it would be most useful to design an app that could show visually the magnitude of proportions and standard deviations being considered in the discussions around the relative severity of different sentence outcomes. By projecting such visualisations in real-time, as the discussion takes place, participants could obtain a more intuitive appreciation of their suggested parameters, which should also contribute to frame the discussions and enhance the consistency of the process. Furthermore, this app could not only help visualise the numerical parameters to be elicited - i.e.1 the input information that will be fed into the Thurstone model, i.e.2 the elicited views in relation to how often a given sentence is thought to be more severe, or the relative range of severity that could be attributed to such sentence - but also how those views will be translated into severity scores - i.e. the output of the Thurstone model.

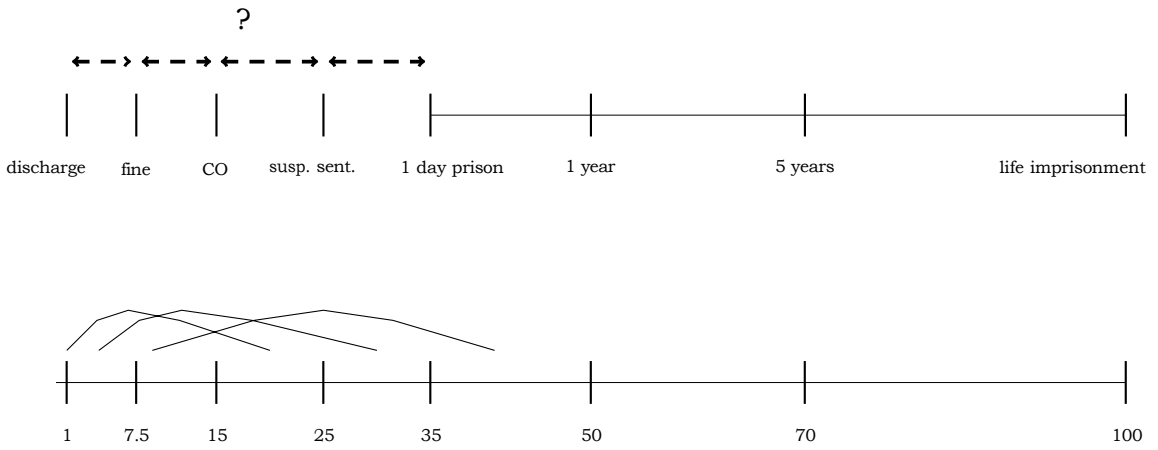


Fig. 2 Conceptual representation of a visual app designed to facilitate the interpretation and discussion of the concepts of penal exchangeability, unequal ranges of severity, and their translation into severity scores, for a selection of sentence outcomes.

Figure 2, shows a conceptual representation of how such visual app could operate, illustrating in different planes the positions of the sentence outcomes being discussed, their potential severity ranges, or their degree of overlap, while simultaneously estimating the position of each outcome in the severity scale according to the information provided. In addition to making some of the numerical elicitations more intuitive, or providing a better framing for the group discussions, being able to see the final product - i.e. the severity scores estimated for each of the sentences considered - would provide an additional layer of robustness to the process. Specifically, participants will be able to assess the face validity of the severity scores to be derived, which would provide them with the opportunity

to reconsider any of the parameters discussed during the workshop. In the event that any issues are raised, the parameter in question could be modified as part of an iterative process, until an agreement is reached and participants are satisfied with the final scale of severity.

6 Conclusion

In this study we have used expert elicitation techniques to explore the validity of a key assumption invoked in studies estimating the relative severity of different sentence types. Namely, we have investigated the assumption of equal variances underpinning the Thurstone scaling method (Thurstone, 1927; Mosteller, 1951; Buchner, 1979; Kwan et al., 2000). An assumption that is rarely stated explicitly, but one we have demonstrated how, when applied to the estimation of sentence severity, is clearly violated. Based on insights elicited from six sentencing experts, we noted wide differences in the range of severity covered by some of the main disposal types used in England and Wales. Our experts agreed that the intervals of severity that could be associated with different types of fines (according to the amount imposed), or custodial sentences (according to their duration), are much narrower than previously considered, while community orders and suspended sentences could encompass much wider severity ranges, depending on the conditions attached to them.

Further, we have demonstrated how accounting for the unequal variances seen across disposal types matters, as the estimated severity scores for specific sentence outcomes using Thurstone scaling vary substantially depending on whether the equal variances assumption is invoked or not. Using the elicited insights into the exchangeability and unequal severity ranges characterising eighteen sentence outcomes, and a modified version of the Thurstone method allowing for unequal variances, we have been able to estimate a new scale of sentence severity.

Ultimately, the goal of the new scale of severity presented here is to be used as an analytical tool to facilitate more robust and insightful quantitative sentencing research. Research that will be better equipped to shed much needed new light on the study of differences across non-custodial outcomes. A highly heterogeneous group of disposal types, representing the vast majority of the sentencing practice, which, for methodological reasons, have remained largely overshadowed by studies focusing on more technically convenient sentence outcomes, such as the probability or duration of custodial sentences.

References

Albonetti CA (1998) The role of gender and departures in the sentencing of defendants convicted of a white-collar offense under the Federal Sentencing Guidelines. Sociology of Crime, Law and Deviance 1:3–48

Allen R (2016) The Sentencing Council for England and Wales: Brake or accelerator on the use of prison? Tech. rep., Transform Justice, URL http://www.transformjustice.org.uk/wp-content/uploads/2016/12/TJ-DEC-8.12.16-1.pdf

Barbora H, Bijleveld C, Smeulers A (2012) Consistency of international sentencing: ICTY and ICTR case study. European Journal of Criminology 9(5):539–552

Baumer EP (2013) Reassessing and redirecting research on race and sentencing. Justice Quarterly 30(2):231–261

Berkson J (1950) Are there two regressions? Journal of the American Statistical Association 45(250):164–180

Black C, Warren R, Ormston R, Cyrus T (2019) Public perceptions of sentencing: National survey report. Tech. rep., Scottish Sentencing Council, URL https://www.scottishsentencingcouncil.org.uk/media/1996/20190902-public-perceptions-of-sentencing-report.pdf

Bolger F (2018) The selection of experts for (probabilistic) expert knowledge elicitation. In: Dias LC, Morton A, Quigley J (eds) Elicitation, pp 393–443

Booth C, Thomas L (2021) An expert elicitation of the effects of low salinity water exposure on bottlenose dolphins. Oceans 2(1):179–192

Brennan A, Pollard D, Coates L, Strong M, Heller S (2017) Expected value of sample information for individual level simulation models to inform stop/go decision making by public research funders: A methodology for the dafneplus diabetes education cluster rct. Value in Health 9(20)

Brunton-Smith I, Pina-Sanchez J, Li G (2020) Re-assessing the consistency of sentencing decisions in cases of assault: Allowing for within court inconsistencies. British Journal of Criminology 60(6):1438–1459

Buchner D (1979) Scale of sentence severity. The Journal of Criminal Law and Criminology 70(2):182–187

Bushway SD, Johnson BD, Slocum LA (2007) Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology. Journal of Quantitative Criminology 23(2):151–178

Dessai S, Bhave A, Birch C, Conway D, Garcia-Carreras L, Gosling J, Mittal N, Stainforth D (2018) Building narratives to characterise uncertainty in regional climate change through expert elicitation. Environmental Research Letters 13(7)

Doob AN, Webster CM (2003) Sentence severity and crime: Accepting the null hypothesis. Crime and Justice 30:143–195

Dr´apal J (2020) Sentencing disparities in the Czech Republic:Empirical evidence from post-communist Europe. European Journal of Criminology 17(2):151–174

Erickson ML, Gibbs JP (1979) On the perceived severity of legal penalties. The Journal of Criminal Law and Criminology 70(1):102–116

Feldmeyer B, Ulmer JT (2011) Racial/ethnic threat and federal sentencing. Journal of Research in Crime and Delinquency 48(2):238–270

Fleetwood J, Radcliffe P, Stevens A (2015) Shorter sentences for drug mules: The early impact of the sentencing guidelines in England and Wales. Drugs: Education, Prevention and Policy 22(5):428–436

Francis B, Soothill K, Humphreys L, Cutajar Bezzina A (2005) Developing measures of severity and frequency of reconviction. Tech. rep., Lancaster University, URL https://eprints.lancs.ac.uk/id/eprint/50137/1/seriousnessreport.pdf

Freiberg A, Fox R (1986) Sentencing structures and sanction hierarchies. Criminal Law Journal 10:216–235

Ginneken E, Hayes D (2017) "just" punishment? offenders' views on the meaning and severity of punishment. Criminology and Criminal Justice 17(1):62–78
Gosling J (2018) Shelf: the Sheffield elicitation framework. In: Dias L, Morton A, Quigley J (eds) Elicitation, Springer, pp 61–93
Gosling J, Hart A, Mouat D, Sabirovic M, Scanlan S, Simmons A (2012) Quantifying experts' uncertainty about the future cost of exotic diseases. Risk Analysis: An International Journal 5(32):881–893
Harlow RE, Darley JM, Robinson PH (1995) The severity of intermediate penal sanctions: A psychological scaling approach for obtaining community perceptions. Journal of Quantitative Criminology 11(1):71–95
Harris L, Walker S (2021) Sentencing Principles, Procedure and Practice 2022. Sweet and Maxwell
Hayes D (2016) Penal impact: Towards a more intersubjective measurement of penal severity'. Oxford Journal of Legal Studies 36(4):724–750
Hayes D (2018) Proximity, pain, and State punishment. Punishment and Society 20(2):253–254
Hester R, Hartman T (2017) Conditional race disparities in criminal sentencing: A test of the liberation hypothesis from a non-guidelines state. Journal of Quantitative Criminology 33:77–100
Hindelang MJ, Dunn CS, Sutton LP, Aumick AL (1975) Sourcebook of criminal justice statistics, 1974. Tech. rep., US National Criminal Justice Information and Statistics Service, URL https://www.ojp.gov/ncjrs/virtual-library/abstracts/sourcebook-criminal-justice-statistics-1974
Irwin-Rogers K, Perry TH (2015) Exploring the impact of sentencing factors on sentencing domestic burglary. In: Roberts JV (ed) Sentencing Guidelines: Exploring Sentencing Practice in England and Wales, Palgrave, Basingstoke, pp 213–239
Isaac A (2020) Investigating the association between an offender's sex and ethnicity and the sentence imposed at the Crown Court for drug offences. Tech. rep., Sentencing Council for England and Wales, URL https://www.sentencingcouncil.org.uk/wp-content/uploads/Sex-and-ethnicity-analysis-final-1.pdf
Isaac A (2021) Estimating the changes in sentencing severity and requirements for prison places associated with the Sentencing Council's guidelines. Tech. rep., Sentencing Council for England and Wales, URL https://www.sentencingcouncil.org.uk/wp-content/uploads/Changes-in-sentencing-severity-and-prison-places-associated-with-SC-guidelines.pdf
Jansen J, Wang H, Holcomb J, Harvin J, Richman J, Avritscher E, Stephens S, Troung V, Marques M, DeSantis S, Yamal J, Pedroza C (2020) Elicitation of prior probability distributions for a proposed Bayesian randomized clinical trial of whole blood for trauma resuscitation. Transfusion 6(3):498–506
Kane E, Minson S (2022) Analysing the impact of being a sole or primary carer for dependent relatives on the sentencing of women in the Crown Court, England and Wales. Criminology and Criminal Justice
King RD, Johnson KR, McGeever K (2010) Demography of the legal profession and racial disparities in sentencing. Law and Society Review 44(1):1–32
Kurlychek MC, Johnson BD (2010) Juvenility and punishment: Sentencing juveniles in adult criminal court. Criminology 48(3):725–758
Kwan YK, Ip WC, Kwan P (2000) A crime index with Thurstone's scaling of crime severit. Journal of Criminal Justice 28(3):237–244
Leclerc C, Tremblay P (2016) Looking at penalty scales: How judicial actors and the general public judge penal severity. Canadian Journal of Criminology and Criminal Justice 58(3):354–384
Lightowlers C, Pina-Sanchez J (2017) Intoxication and assault: an analysis of Crown Court sentencing practices in England and Wales. The British Journal of Criminology 58(1):132–154
Lightowlers C, Pina-Sanchez J, Watkins E (2020) Contextual culpability: How drinking and social context impact upon sentencing of violence. Criminology and Criminal Justice
Lovegrove A (2001) Sanctions and severity: To the demise of von Hirsch and Wasik's sanction hierarchy. The Howard Journal of Crime and Justice 40(2):126–144

McClelland KM, Alpert GP (1985) Factor analysis applied to magnitude estimates of punishment seriousness: Patterns of individual differences. Journal of Quantitative Criminology 1(3):307—-318
McDavid JC, Stipack B (1981) Simultaneous scaling of offense seriousness and sentence severity through canonical correlation analysis. Law and Society Review 16(1):147–162
Ministry of Justice (2021) Data first: An introductory user guide. Tech. rep., Ministry of Justice, URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/984510/data-first-user-guide-v5.pdf
Mosteller F (1951) Remarks on the method of paired comparisons: I. the least squares solution assuming equal standard deviations and equal correlations. Psychometrika 16(1):3–9
O'Hagan A (2019) Expert knowledge elicitation: Subjective but scientific. The American Statistician 73(1):69–81
O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, Oakley JE, Rakow T (2006) Uncertain Judgements: Eliciting Expert Probabilities. John Wiley, Chichester
Padfield N (2011) Time to bury the "custody threshold"? Criminal Law Review 8:593–612
Petersilia J, Deschesnes E (1994) Perceptions of punishment: Inmates and staff rank the severity of prison versus intermediate sanctions. Prison Journal 74(3):306–328
Pina-S´anchez J, Gosling JP (2020) Tackling selection bias in sentencing data analysis: a new approach based on a scale of severity. Quality and Quantity
Pina-S´anchez J, Linacre R (2014) Enhancing consistency in sentencing: Exploring the effects of guidelines in England and Wales. Journal of Quantitative Criminology 30(4):731–748
Pina-S´anchez J, Lightowlers C, Roberts JV (2016) Exploring the punitive surge: Crown Court sentencing practices before and after the 2011 English riots. Criminology and Criminal Justice 17(3):319–339
Pina-S´anchez J, Gosling JP, Chung H, Geneletti S, Bourgeois E, Marder I (2019a) Have the England and Wales guidelines influenced sentence severity? An empirical analysis using a scale of sentence severity and time-series analyses. British Journal of Criminology
Pina-S´anchez J, Roberts JV, Sferopoulos D (2019b) Does the Crown Court discriminate against Muslim-named offenders? A novel investigation based on text mining techniques. British Journal of Criminology 59(3):718–736
Revelle W (2018) psych: Procedures for personality and psychological research,. Tech. rep., Northwestern University, URL https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research
Roberts JV, Bild J (2021) Ethnicity and custodial sentencing in England and Wales: A review of trends, 2009-2019. Tech. rep., Sentencing Academy, URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3881930
Roberts JV, Pina-Sanchez J (2021) Sentence reductions for a guilty plea: The impact of the revised guideline on rates of pleas and 'cracked' trials. Journal of Criminal Law
Roberts JV, Bild J, Pina-Sanchez J, Hough M (2022) Public knowledge of sentencing practice and trends. Tech. rep., Sentencing Academy, URL https://sentencingacademy.org.uk/wp-content/uploads/2022/01/Public-Knowledge-of-Sentencing-Practice-and-Trends.pdf
Sebba L, Nathan G (1984) Further explorations in the scaling of penalties. The British Journal of Criminology 24(3):221–249
Sentencing Council (2021) Background quality report. Tech. rep., Sentencing Council for England and Wales, URL https://www.sentencingcouncil.org.uk/wp-content/uploads/Background-Quality-Report-Theft-from-a-shop-or-stall-data.pdf
Spelman W (1995) The severity of intermediate sanctions. Journal of Research in Crime and Delinquency 32(2):107–135
Steffensmeier D, DeMuth S (2001) Ethnicity and judges' sentencing decisions: Hispanic-Black-White comparisons. Criminology 39:145—-176
Thomas KJ, Hamilton BC, Loughran TA (2018) Testing the transitivity of reported risk perceptions: Evidence of coherent arbitrariness. Criminology 56(1):59–86

Thurstone LL (1927) A law of comparative judgement. Psychological Review 34:273–286
Tremblay P (1988) On penal metrics. Journal of Quantitative Criminology 4(3):225–245
Ulmer J, Light MT, Kramer J (2011) The “liberation” of federal judges’ discretion in the wake of the Booker/Fanfan decision: Is there increased disparity and divergence between courts? Justice Quarterly 28(6):799–837
Ulmer JT, Eisenstein J, Johnson BD (2010) Trial penalties in federal sentencing: extra-guidelines factors and district variation. Justice Quarterly 27(4):560–592
Usher W, Strachan N (2013) An expert elicitation of climate, energy and economic uncertainties. Energy Policy 61:811–821
Yan S, Lao J (2021) Sex disparities in sentencing and judges’ beliefs: A vignette approach. Victims and Offenders

Technical Appendix: R Code

```
################################
# Thurstone - Unequal Variances #
################################

###### Enter number of outcomes under consideration
num_outcomes = 18

###### Enter customised standard deviations
latent_sds = sqrt(1/2) * c(0.01, 0.625, 0.375, 0.375, 0.375, 0.375, 0.375, 0.375,
                1, 1, 1.25, 1, 1.25, 1.25, 1.75, 0.25, 0.25, 0.25)

###### Enter matrix of preferences
elicited_proportions = matrix(c(
0.5,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,
0.001,0.5,0.65,0.8,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,
0.001,0.35,0.5,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,
0.001,0.2,0.001,0.5,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,
0.001,0.001,0.001,0.001,0.5,0.999,0.999,0.999,0.65,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,
0.001,0.001,0.001,0.001,0.001,0.5,0.999,0.999,0.55,0.75,0.999,0.75,0.999,0.999,0.999,0.999,0.999,0.999,
0.001,0.001,0.001,0.001,0.001,0.001,0.5,0.999,0.4,0.6,0.95,0.7,0.75,0.999,0.999,0.999,0.999,0.999,
0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.5,0.2,0.5,0.9,0.6,0.65,0.8,0.999,0.999,0.999,0.999,
0.001,0.001,0.001,0.001,0.001,0.45,0.6,0.8,0.5,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,0.999,
0.001,0.001,0.001,0.001,0.001,0.25,0.4,0.5,0.001,0.5,0.999,0.55,0.999,0.999,0.999,0.999,0.999,0.999,
0.001,0.001,0.001,0.001,0.001,0.001,0.05,0.1,0.001,0.001,0.5,0.4,0.45,0.6,0.999,0.6,0.7,0.75,
0.001,0.001,0.001,0.001,0.001,0.25,0.3,0.4,0.001,0.45,0.6,0.5,0.999,0.999,0.999,0.999,0.999,0.999,
0.001,0.001,0.001,0.001,0.001,0.001,0.25,0.35,0.001,0.001,0.55,0.001,0.5,0.95,0.999,0.999,0.999,0.999,
0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.2,0.001,0.001,0.4,0.001,0.05,0.5,0.999,0.85,0.999,0.999,
0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.5,0.3,0.35,0.4,
0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.4,0.001,0.001,0.15,0.7,0.5,0.999,0.999,
0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.3,0.001,0.001,0.001,0.65,0.001,0.5,0.999,
0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.001,0.25,0.001,0.001,0.001,0.6,0.001,0.001,0.5),
num_outcomes, num_outcomes, byrow=TRUE)

###### Function for creating preference matrix from latent parameters
create_matrix = function(latent_means){
# Empty matrix to store
pref_probs = matrix(0, num_outcomes, num_outcomes)
# Calculate probability of one latent random variable being greater than another
for (i in 1:num_outcomes){
for (j in 1:num_outcomes){
pref_probs[i,j] = pnorm(0,
latent_means[i] - latent_means[j],
sd = sqrt(latent_sds[i]^2 + latent_sds[j]^2))
}
}
# Return matrix
return(pref_probs)
}

###### Function for measuring distance between created matrix and elicited proportions
matrix_distance = function(latent_means){
# set first latent mean to be 0
latent_means = c(0,latent_means)
# return distance based on Frobenius norm
return(Matrix::norm(create_matrix(latent_means)-elicited_proportions, type = 'F'))
}

###### Find the latent means
optim_out <- optim(1:(num_outcomes-1), # We are trying to find num_outcomes - 1 because the first is fixed at 0
```

```
matrix_distance, method = 'L-BFGS-B', lower = 0, control = list())
# Has it converged?
ifelse(optim_out$convergence == 0, TRUE, FALSE)
# Found latent means
c(0,optim_out$par)
# Fitted preference matrix
round(create_matrix(c(0,optim_out$par)),3)
# Original matrix
elicited_proportions


#####################
# Custodial scores  #
#####################

#Immediate custody
month = c(1, 2, 3)
sev = optim_out$par[15:17]
#A linear function to estimate severity of immediate custody > 3 months
linear_imm = lm(sev ~ month)
summary(linear_imm)
#We can now predict severity scores for different immediate custody sentences using the above function, i.e.:
#severity = 12.54 + 0.43*custody_months
#Predicted score for a 12 months immediate custody
linear_imm$coefficients[1] + 12*linear_imm$coefficients[2]
#Predicted score for a 24 months immediate custody
linear_imm$coefficients[1] + 24*linear_imm$coefficients[2]
#Predicted score for a 60 months immediate custody
linear_imm$coefficients[1] + 60*linear_imm$coefficients[2]
#Predicted score for a 120 months immediate custody
linear_imm$coefficients[1] + 120*linear_imm$coefficients[2]
#Predicted score for a 240 months immediate custody
linear_imm$coefficients[1] + 240*linear_imm$coefficients[2]

#Suspended sentences
month_cust = c(1, 1, 6, 12)
month_susp = c(6, 12, 6, 24)
sev = optim_out$par[11:14]
#A linear function to estimate severity of immediate custody > 3 months
linear_susp = lm(sev ~ month_cust + month_susp)
summary(linear_susp)
#We can now predict severity socres for different suspended sentences using the above function, i.e.:
#severity = 10.73 + 0.20*custody_months + 0.04*suspended_months
#3 months custody suspended for 12 months - predicted score
linear_susp$coefficients[1] + 3*linear_susp$coefficients[2] + 12*linear_susp$coefficients[3]
#9 months custody suspended for 12 months - predicted score
linear_susp$coefficients[1] + 9*linear_susp$coefficients[2] + 12*linear_susp$coefficients[3]
#9 months custody suspended for 24 months - predicted score
linear_susp$coefficients[1] + 9*linear_susp$coefficients[2] + 24*linear_susp$coefficients[3]


##########################################
# Constraining the scale to a 0-100 range #
##########################################

#First we need to contemplate the maximum severity score possible
#Here we have assumed that is a 20 years custodial sentence, which was estimated at 115.56 severity
#Then we need to re-scale accordingly by multiplying severity scores by the following: 100/115.56
#So, a low community order with severity equal to 8.066 will be rescaled as:
4.86 * 100/115.56

#A table with the severity scores for the main sentence outcomes considered could be presented as follows:
```

```
outcomes = c("absolute discharge", "conditional discharge", "fine-A", "fine-B", "fine-C", "fine-D",
"fine-E", "fine-F", "community order-low", "community order-medium", "community order-high",
"1-month custody suspended for 6 months", "1-month custody suspended for 12 months",
"6-month custody suspended for 6 months", "12-month custody suspended for 24 months",
"1-month immediate custody", "2-month immediate custody", "3-month immediate custody",
"12-month immediate custody", "60-month immediate custody", "120-month immediate custody",
"240-month immediate custody")
severity = c(0, optim_out$par[1:17],
linear_imm$coefficients[1] + 12*linear_imm$coefficients[2],
linear_imm$coefficients[1] + 60*linear_imm$coefficients[2],
linear_imm$coefficients[1] + 120*linear_imm$coefficients[2],
linear_imm$coefficients[1] + 240*linear_imm$coefficients[2])
severity_rescaled = severity * 100/115.56
cbind(outcomes, severity_rescaled)
```

From Legislative Intent to Hospice Practice: Exploring the Genealogy of the Mental Capacity Act 2005

*Dr Caroline. A. B. Redhead**

Abstract

The Mental Capacity Act 2005 (MCA) regulates decision-making for people without capacity. Post-legislative scrutiny of the Act in 2014 by a House of Lords Select Committee concluded that the MCA was neither well understood nor working well in practice. The aim of the research discussed in this article was to consider how the Act's principles are understood and interpreted in hospice practice, specifically considering the patient's role in the decision-making process.

The research proceeded through four distinct, but linked, phases which, together, offered a 'life story' of the MCA from legislative intent to current hospice practice (in 2019). The research was informed by relational theory and legal consciousness theory and the methods described are underpinned by a narrative approach to analysis. Phase one was an innovative genealogical analysis of policy and legislative documents (n=24) influencing the 'coming to be' of the MCA. In phase two, a systematic review of Court of Protection judgments (n=63) 'historicises' the empirical research, which was the focus of phases three and four (group interviews and individual interviews, respectively). Staff from two participating hospices participated in two group interviews and six individual interviews (13 participants), providing empirical data. Template analysis was used in all four phases of the study, and adapted to facilitate a synthesis of the findings across the study as a whole.

1. Introduction

The way we care for people approaching the end of life, and how we manage death, are continuing conversations that the hospice community in the UK wants to encourage.[1] These conversations often take place as hospice patients lose their ability to make decisions for themselves as the end of their life approaches, creating a unique decision-making context. An understanding of how patients' decision-making should be properly supported (both legally and as part of their care) is thus particularly important for hospice staff. In England and Wales, the Mental Capacity Act 2005 (MCA) sets out what constitutes capacity to make a decision, how to assess whether someone has capacity, and how decisions can be made for individuals who lack the capacity to make them for themselves. The Act came into force in 2007 and was intended to catalyse and lead societal change.[2] It has been described as a 'visionary piece of legislation', which 'marked a turning point in the statutory rights of people who may lack capacity' because it 'place[d] the individual at the heart of decision-making'.[3] Yet, post-legislative scrutiny of the MCA by a House of Lords Select Committee in 2014 concluded that the Act was not working well in practice, that its implementation had not delivered the empowerment it promised, and that cultures of paternalism (in health) and risk aversion (in

* School of Law, Centre for Social Ethics and Policy, The University of Manchester, Oxford Rd, Manchester, M13 9PL, UK. The authors is grateful to Professor Sara Fovargue, Dr Lucy Frith and Professor Marie Fox for providing comments on a draft of this paper, and to the anonymous reviewers of the submitted draft for their helpful suggestions.

[1] Hospice UK <www.hospiceuk.org/our-campaigns/dying-matters> accessed 16 May2022.
[2] Select Committee on the Mental Capacity Act 2005*, Mental Capacity Act 2005: post-legislative scrutiny* (HL 2013-14, HL 39); Department for Constitutional Affairs, *Mental Capacity Act 2005 Code of Practice* (The Stationary Office, 2007). Available at <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/921428/Mental-capacity-act-code-of-practice.pdf> accessed 16 May 2022.
[3] Select Committee on the Mental Capacity Act 2005 (n 2) 6.

social care) continued to prevail.[4] The Select Committee considered that, in the context of health and social care, the MCA was not widely understood and embedded in practice.[5]

I explored whether these conclusions were accurate for the understanding and implementation of the MCA in English hospices, by investigating how hospice staff cared for patients whose ability to make decisions for themselves had been compromised by the progression of their disease. In this article I describe the qualitative methods I developed and used in my research to map the key ideas of the policy-makers and legislators who were engaged in the development of the MCA and then trace them through decisions in the Court of Protection, and into policy and practice in hospices. My aim was to construct a genealogy of the MCA from legislative intent to (then) current hospice practice, to chart the dynamic relationship between law and practice in the area of decision-making for people without capacity, and, ultimately, to compare the decision-making approach of hospice staff to the intentions of the original legislators. My conclusion, challenging the findings of the Select Committee in the particular context of hospices, was that hospice staff *did* understand the MCA and were implementing it in hospice practice to the benefit of those whose decision-making capacity had been compromised by the progression of their disease.

In this article I present both the methodology and the methods I adopted to guide my enquiry. In section 1, I introduce my research and set out the context for it, I note the theoretical framing in section 2, and describe and discuss the methods and the methodological approach in sections 3-6. In the concluding section of the article, I reflect on the methodological choices, how they supported my investigation, and whether this combination of methods has broader application to empirical socio-legal analysis.

[4] Select Committee on the Mental Capacity Act 2005 (n 2) 7
[5] Select Committee on the Mental Capacity Act 2005 (n 2) 8

2. Setting the Scene: An Overview of the Study and the Landscape of the Article

My methodological approach evolved iteratively, in tandem with the design of my empirical study, as I began to engage with literatures of interest. I wanted to present an interpretation of the 'life story' of the MCA from conception to practice in the hospice context, aiming to offer a critical exploration of *how* the Act was working in practice to (re)assess the Select Committee's conclusions. I used four phases of enquiry to facilitate the exploration of different, but complementary, data sources. In this way, I constructed a layered and comprehensive socio-legal account of the MCA's approach to decision-making for people who lack the capacity to make decisions for themselves.

In the first phase, my focus was on documentary sources that illuminated the social and policy context within which the concepts underpinning the MCA were debated, and its principled approach to decision-making established.[6] Examining this contemporary discourse gave me a window into the legal and ethical context from which the MCA emerged. In phase two, I designed and carried out a systematic review of decisions in the Court of Protection (CoP), to underpin an analysis of the way in which the courts had interpreted the key concepts of the MCA since its entry into effect. In phases three and four, I undertook empirical research, narrowing my focus to the hospice context and investigating organisational practice (phase three), and staff members' individual experiences of decision-making with patients (in phase four).

In designing this phased approach, I was conscious that, for the successful completion of the research as a whole, a coherent transition between each phase would be essential. The theoretical framing and methodological approach had to 'hold' all four phases of the study which, albeit all qualitative in nature, would draw data from a variety of documentary and

[6] Noting that the 'principled' approach had not been drafted into any piece of legislation in England and Wales before the MCA. See Joint Committee on the Draft Mental Incapacity Bill, *Draft Mental Incapacity Bill* (HL Paper 198-II, HC 1083-II, TSO 2003), particularly Vol I Recommendations 4 and 5

empirical sources. My objective was to analyse data from each distinct phase using a method that could identify and preserve any themes flowing between them; thus allowing the research to reflect the developing story of the MCA. I considered a number of analytic methods which would both enable this flow and allow me to achieve a coherent synthesis without losing the granularity of the data from each phase. Common to each phase, and supporting the coherence of the research as a whole, was the embedding of the data in the language and concepts that describe and situate the MCA. For this reason, I will now briefly describe the contours of the MCA decision-making landscape and introduce the key principles and concepts (capacity and best interests) which both guide and constrain decision-makers. Equally important is the context of the hospice 'movement' in the UK, and I next explain its salient features, so that the key themes which link the phases of the research and the methodological relationships between them in the (arguably) unique context of hospice care are situated.

The MCA, 'capacity' and 'best interests'

The MCA provides the legal framework for decision-making on behalf of individuals who lack the capacity to make decisions for themselves. The Act aims to protect people who lacked the capacity to make a decision, while empowering them by maximising their ability to decide or to participate in the decision-making as far as they were able.[7] The MCA is underpinned by five key principles:

1. A person must be assumed to have capacity unless it is established that they lack it (s1(2)).
2. A person must not be treated as unable to make a decision unless all practicable steps to help them to do so have been taken without success (s1(3)).

[7] Department for Constitutional Affairs (n 2).

3. A person must not be treated as unable to make a decision merely because they make an unwise decision (s1(4)).
4. An act done, or decision made, under the MCA for or on behalf of a person who lacks capacity must be done, or made, in their best interests (s1(5)).
5. Before the act is done, or the decision is made, regard must be had to whether the outcome can be as effectively achieved in a way that is less restrictive of the person's rights and freedoms of action (s1(6)).

Capacity may fluctuate and an individual may have capacity to make some decisions but not others, the key is whether 'at the material time' (MCA s2(1)) someone can make *and* communicate the decision, whether by talking, using sign language or any other means (MCA s3(1)). To have capacity to make a decision, an individual must be capable of understanding what is proposed, of retaining, using and weighing information in the process of making the decision, and of communicating that decision (MCA s3). If they cannot make (or communicate) a decision because of an impairment of, or a disturbance in the functioning of, the mind or brain, capacity will not be established (MCA, s2(1)).

In the event that someone is considered *not* to have capacity to make the decision for themselves, the MCA permits a decision to be made on their behalf, in their best interests (MCA s1(5)). A range of factors must be considered in establishing someone's best interests, including their past and present wishes and feelings, the beliefs and values likely to influence their decision and any other factors they would consider if they were able to (MCA s4(6)). The MCA does not prioritise a person's wishes and feelings above any other of the relevant factors. The lack of priority given to someone's wishes and feelings in the MCA framework has been the

subject of academic criticism,[8] and goes against the approach set out in the United Nations Convention on the Rights of Persons with Disabilities.[9] Indeed the Law Commission, in its recent review of the MCA, recommended that particular weight *should* be accorded to someone's wishes and feelings in making a decision for or about them.[10] Despite the Government's decision not to legislate to reflect this, it is clear that the wishes and feelings of the person about whom a decision is being made are an increasingly important factor in a 'best interest' decision about them[11]. Hospices also espouse and promote a 'patient-centred' approach in their provision of palliative care,[12] and I wanted to investigate whether attention to patients' wishes and feelings was a characteristic of this approach in the context of MCA decision-making by hospice staff. I therefore paid particular attention to exploring the ways in which patients were supported to make decisions, and the extent to which their wishes and feelings were reflected in hospice decision-making.

The 'hospice movement' in the UK

The work of Cicely Saunders, particularly her wish to avoid both neglect of the dying and the medicalisation of death, is credited with giving rise to what has become known as 'the hospice movement', which was created to prioritise care for patients with a terminal diagnosis and to promote an holistic approach to their care.[13] Saunders' concept of 'total pain' underpins hospice philosophy, which is that, in providing palliative care, hospices must acknowledge and

[8] See, for example, Emily Jackson, 'From 'Doctor Knows Best' to Dignity: Placing Adults Who Lack Capacity at the Centre of Decisions About Their Medical Treatment' (2018) 81 *The Modern Law Review* 2 247.
[9] Convention on the Rights of Persons with Disabilities (adopted 13 December 2006, entered into force 3 May 2008) 2515 UNTS 3 (CRPD) art 3.
[10] Law Commission *Mental Capacity and Deprivation of Liberty; A Consultation Paper* (Law Com No 222 2015). Available at <www.lawcom.gov.uk/project/mental-capacity-and-deprivation-of-liberty/> accessed 16 May 2022. Note that, while these recommendations were accepted by the Government, the Mental Capacity (Amendment) Act 2019 (which amends the MCA and came into force in May 2019) does not reflect them.
[11] *Wye Valley NHS Trust v B* [2015] EWCOP 60.
[12] Hospice UK <www.hospiceuk.org/information-and-support/your-guide-hospice-and-end-life-care/im-looking-hospice-care/what-hospice#content-menu-1780> accessed 16 May 2022.
[13] Cecily Saunders, 'The evolution of palliative care' (2000) 41 Patient Education and Counselling 7; David Clark *Cicely Saunders: A Life and Legacy* (Oxford University Press, 2018).

address not only patients' physical pain, but also their emotional and social pain, and their spiritual need for security, meaning and self-worth.

The notion of person-centred care, described by hospices as fundamental to their approach, is undergirded by the concept of 'total pain.' These two key concepts come together in the over-arching aim of the hospice movement to tailor palliative care around the needs of *that* particular patient and *their* particular condition.[14] I suggest that this approach, together with the distinctive evolution of UK hospices outside the NHS[15] and their aim to support (but not to cure) patients at the end of life, means that hospices interpret the MCA in a unique decision-making context. In situating my empirical work in hospices, I was interested in exploring how this unique context was reflected in, or translated into, hospice staff members' interpretation of the key principles of the MCA, particularly the concepts of 'capacity' and 'best interests'.

What I found was that the empirical data, grounded in the hospice as a particular decision-making context, encouraged a focus on the importance of thinking relationally. Interview participants suggested that relationships become more widely relevant at the end of life, so that, for example, their patients' relationships with past selves, with their current disease, and with an imagined future all influenced decision-making. Decisions were taken within the context of a multi-disciplinary team, of which the patient and carers or family members were an integral part. This encouraged a methodological approach, particularly to the synthesis of the four phases, that was similarly attentive to the workings of relational entanglements.

3. Theoretical Framing and Methodology

[14] Hospice UK (n 12).
[15] Hospices in the UK are typically constituted as independent charities, see Hospice UK <https://professionals.hospiceuk.org/about-us/membership> accessed 2 September 2022.

Acknowledging the centrality of relationships to MCA decision-making in hospices, my research investigated relational theories in the context of the MCA framework. In my exploration of the 'constellation of ideas, practices and institutions'[16] relevant to the notions of self and autonomy that inform legal ideas about capacity and decision-making, I was influenced by the ideas of relational theorists, whose convictions I share. Relational theory posits that, 'existence is not an individual affair,[17]' but, rather, that we are embedded in, and constituted by, our relationships.[18] This theoretical framing, in the context of hospice care, allowed for an exploration of the law's approach to care and compassion as relational concepts,[19] and, as part of that, an engagement with ethic of care literature. I found Carol Gilligan's understanding of an ethic of care as a guide to acting carefully, understanding the costs of not paying attention, not listening, of being absent rather than present and of not responding to another with integrity and respect particularly relevant to MCA decision-making.[20] My data indicated that an approach akin to this guides hospice staff in their interpretation of the MCA for patients approaching the end of life. I also drew on legal consciousness theory[21] to explore the way in which the MCA is experienced by hospice staff members as they make decisions in the context of caring for patients.

Interlinked with my relational approach to the law, 'story' was another key feature of my research. I thought about stories in my genealogical consideration of the 'life story' of the

[16] Jennifer Nedelsky, *Law's Relations A Relational Theory of Self, Autonomy, and Law* (OUP 2011) 3.
[17] Karen Barad, *Meeting the Universe Halfway: quantum physics and the entanglement of matter and meaning* (Duke University Press 2007) ix.
[18] Carol Gilligan, 'In a Different Voice: Women's Conception of the Self and of Morality' (1977) 47 Harvard Educational Review 4 481; Barad, (n 17); Jonathan Herring, 'Caring and the law' (Hart 2013); Nedelsky (n 16).
[19] Maksymilian Del Mar, 'Imagining by feeling: a case for compassion in legal reasoning' (2017) 13 International Journal of Law in Context 2 143.
[20] Carol Gilligan, 'Moral Injury and the Ethic of Care: Reframing the Conversation about Differences' (2014) 45 Journal of Social Philosophy 1 89. I was also influenced by Gilligan's earlier work (see n 18), the work of Jonathan Herring, see n 18 and 'Compassion, ethics of care and legal rights' (2018) 13 International Journal of Law in Context 2 158; and the work of Joan Tronto, see, for example, 'Beyond gender difference to a theory of care' (1987) 12 Signs: Journal of Women in Culture and Society 4 644.
[21] Patricia Ewick and Susan Silbey, *The Common Place of Law: stories from everyday life* (University of Chicago Press 1998).

MCA and my characterisation of judgments as legal stories linking the 'law in books' to the law in ordinary people's everyday narratives of life and circumstances. My empirical work engaged with stories in the more traditional sense, attending to the stories told by hospice staff members to illustrate their understanding and interpretation of the MCA in practice. A narrative approach is particularly relevant to hospice decision-making, because stories have long been 'given time and space' in hospice and palliative care.[22] Furthermore, in the specific context of an MCA decision-making process, the patient's story is a key aspect of a proper consideration of their best interests.[23]

A narrative approach also suited the non-linear view of time that underpinned my research. Stories contract time, wander back and forth in time, are inextricably part of time and yet remain timeless as they are re-made each time they are (re)told and actively reconfigure the past within the context of their telling.[24] Narrative research can, therefore, use storied tellings to investigate how the past is brought to bear, how cultural memory (including legal memory) is involved in historical narratives, and how the present is constituted from stories about the past, including the processes and procedures of law.[25] Adopting a storied approach to thinking about data in each of the four phases also created conceptual and analytic coherence. I followed Reissman's approach, treating archival documents, judgments and interviews as analytic units (thus keeping the narrative intact) rather than fragmenting them into coding units for analysis.[26] This allowed me not only to achieve a more nuanced analysis of the content of each *individual* document, but also to identify patterns (themes) chronologically across each *phase* of the research and, ultimately, across the study as a whole. I was also persuaded by Reissman's

[22] Amanda Bingley et al, 'Developing narrative research in supportive and palliative care: the focus on illness narratives' (2008) 22 Palliative Medicine 653.
[23] See, for example, Warren Brookbanks, *Narrative Medical Competence and Therapeutic Jurisprudence: Moving towards a synthesis* (in McMahon, M. & Wexler, D. (eds) *Therapeutic Jurisprudence,* The Federation Press 2003).
[24] Mona Livholts and Maria Tamboukou *Discourse and narrative methods* (Sage 2015).
[25] Ewick and Silbey (n 21).
[26] Cathy Reissman, *Narrative methods for the human sciences* (Sage 2008).

theoretical conviction that narrative is an important approach to investigative enquiry because a close analysis of stories can reveal truths about human experience.[27]

My contention is that law is part of this narrative connection between people and society.[28] Law does not sit outside the social world in a separate, distinct, rational and objective reality. Rather, law is experienced as part of society, woven through individuals' private and professional relationships.[29] This theoretical and methodological approach underpinned and informed the methods I developed and will discuss in this article, starting with my exploration of the social and policy context within which the concepts underpinning the MCA were debated.

Foucauldian Genealogy as Method: Analysis of the Descent and Emergence of the MCA

Foucauldian genealogy focuses on the descent of the subject; descent in the sense of a lineage, a family tree or a network of relationships.[30] Foucault did not consider that anything was traceable back to a single point of origin; rather, he looked at heterogeneous and diverse historical sources and explored the links between them in order to understand how they worked together to set the scene for something new to come about.[31] Foucault was interested in relationships with the power to effect change and to establish new ways of thinking, and in the ways in which the interplay of these relationships promoted new possibilities and ways of being in the world.[32] In Foucauldian genealogical analysis, the aim is to trace the antecedents of the subject of interest, to look for factors that might have come together to set the scene (to create the conditions of possibility) the start of something new.[33]

---

[27] Reissman ibid.
[28] For a discussion of the idea of law as a biography, see Sally Sheldon and others, 'The Abortion Act (1967): a Biography (2019) 39 Legal Studies 18.
[29] Ewick and Silbey (n 21); Nedelsky (n 16).
[30] C G Prado, *Starting with Foucault: an introduction to genealogy* (Westview Press 2000).
[31] Ibid.
[32] Ibid.
[33] Michel Foucault, 'Nietzsche, genealogy and history'. (Paul Rainbow (ed) *The Foucault Reader* Penguin 1975).

In designing a method informed by Foucauldian thinking, my aim was to untangle and explore the various threads and connections which came together at that point in history when the adequacy of the law concerning mental capacity and decision-making was questioned.[34] In looking to trace the descent of the MCA, I sought a broad understanding of the direction of *social* policy travel, in order to investigate its relevance to *legal* change. I aimed to create a data archive that 'took the temperature' of society immediately prior to the emergence of the MCA, and to use that archive as a tool with which to bring Foucault's genealogical approach to bear in the socio-legal context. In the absence of any established Foucauldian method for undertaking a socio-legal genealogical analysis,[35] I designed and implemented a two-stage approach. In the first stage (*Descent*), I constructed a data archive. In the second stage (*Emergence*), after reading and re-reading the data in the archive, I looked for patterns (themes[36]), insights into the changes in societal thinking, that became influential in the changed societal and legislative approach to mental incapacity.

It was in the first stage, 'descent' that the key challenge of this genealogical approach lay. In compiling the data archive, I was creating an imaginary of the 'descent' of the MCA's approach and key concepts, the first chapter in the 'life story' I wanted to write. I was conscious that an insufficiently broad-ranging archive might omit aspects of the story, limiting the patterns and insights I was hoping to reveal. To guard against this, I looked for threads and connections widely in a diverse collection of documentary materials. I continued building the archive until I felt that I had achieved a level of data saturation, that my continued reading was confirming similarities in the data rather than new 'lines of flight,' thus creating a productive symbiotic relationship between the 'descent' stage and the 'emergence' stage.

[34] See Lucy Series, *Deprivation of Liberty in the Shadows of the Institution,* (Bristol University Press 2022) for a similar 'critical genealogical' approach to investigate problems associated with social care.
[35] Prado (n 30).
[36] These informed my use of template analysis to explore the data, see Section 6 below.

My exploration of the historical material to map the descent of the MCA started with the documents cited in the explanatory note to the Act.[37] I adopted a 'snowballing' approach and selected additional documents for review as I identified connections and tessellations. This approach generated a heterogeneic archive of historical documents which would have been difficult to access by way of a traditional review of a particular body of literature. The 'snowballing' approach, by contrast, allowed me to search across a broad variety of documents, and to create an archive comprising legal case reports, consultation documents, policy documents, judicial opinion, parliamentary reports, UN documents, press reports and political manifestos. My construction of the archive was deliberately neither linear nor chronological. I followed the references and connections backwards and forwards in time, identifying the factors that came together to set the scene for a legislative change and created the 'conditions of possibility' for the emergence of the MCA; in particular, the key principles (in s1 of the MCA) and the central concepts of capacity and best interests, which were intended (and continue) to direct the Act's use in practice.

These concepts and principles became visible as far back as 1991[38], and the philosophy underlying them was visible throughout the archive as a whole. Immersing myself in the documents, I reflected on whether, and how, they influenced the development of the MCA's s1 principles and central concepts. By interrogating those reflections, I identified three threads running through the archive: an increasing societal acceptance (and celebration) of difference, a movement towards inclusion and empowerment, and an intolerance of discrimination or misuse of power. I drew three key over-arching themes from these threads; *emergence of the individual*, *a person of value*, and *role of law*. There was support across the archive for legislation which enabled rather than restricting, and which accorded rights to a person without

[37] Available at < www.legislation.gov.uk/ukpga/2005/9/notes/contents> accessed 16 May 2022.
[38] Law Commission, *Mentally Incapacitated Adults and Decision-Making: An Overview* (Law Com No 119 1991).

capacity (the *emergence of the individual)*. The *person of value* theme references the philosophical idea that human beings are valuable in and of themselves.[39] This underpins the human rights instruments that appeared in my data archive, and, in particular, the importance of their emphasis on the dignity and worth of each human being.[40] The entry into effect of the Human Rights Act 1998 and the Disability Discrimination Act 1995 during the period when the MCA was being discussed reflects the influence of a human rights discourse on policy during this period. The *role of law* theme, central to my overall interest in the relationship between law and society, developed throughout the study (leading to my use of legal consciousness theory in understanding the experience of hospice staff members in applying the MCA framework). In the context of my Foucauldian method, the *role of law* recognizes Foucault's characterisation of legal (and medical) practices and processes of 'legality' as powerful.[41] His conception of power as a complex network, within which people act to condition the options and actions of others, also has particular resonance in the context of medical decision-making for people without capacity.[42]

Working together through the phases of the research, then, these three themes encapsulate the social change that took place in the decade prior to the enactment of the MCA. They became my key themes, and, linking the four phases of the research, they underpinned my detailed investigation of the development of the law concerning mental capacity from the intentions of those debating law reform in the 1990s to the interpretation of the MCA in English hospices in the present day. The next stage in that investigative journey was my review of Supreme Court and CoP judgments from 2007, when the MCA came into force, until 2018.

---

[39] Tom Beauchamp and James Childress, *Principles of Biomedical Ethics* (8th edn, OUP 2009).
[40] See, for example, United Nations *Declaration on the Rights of Mentally Retarded Persons,* <ohchr.org/EN/ProfessionalInterest/Pages/RightsOfMentallyRetardedPersons.aspx> accessed 16 May 2022.
[41] Michel Foucault, *Discipline and punish: The birth of the prison* (A. Sheridan tr, Penguin 1975).
[42] Ibid.

4. The Embedding of the MCA into Practice

The effect of law depends, amongst other things, on how it is interpreted and applied by the courts. Judicial action sits alongside political, institutional and social conceptions and understandings of law.[43] As propositions of law, judgments combine elements of both description and evaluation, and, thus, become interpretive of legal history as it applies in the present.[44] Judicial decisions have direct consequences on the individuals involved in the cases, and indirect, longer-term consequences as those decisions are applied and (re)interpreted in subsequent cases heard by the courts. I used Supreme Court and CoP judgments as lenses through which to review how the MCA had been, and was being, embedded in practice.

I adapted the requirements for a systematic literature review in the social science tradition to my search of case law interpreting and applying the MCA, designing and using a systematic strategy to ensure that I completed a comprehensive search for relevant judgments. Case law applying the MCA is, unsurprisingly, extensive, and so I applied specific inclusion and exclusion criteria to enable the identification of an appropriately focused body of judgments for detailed analysis. My search was directed by this question:

How have the courts in England and Wales interpreted the meaning of 'capacity' and 'best interests', including the role of an individual, supporting family members and carers (formal and informal), when decisions are made under the MCA about medical care and treatment?

I included or excluded cases by applying the following criteria**:**

Inclusion: cases considering:

[43] Ewick and Silby (n 21); C. Smulavitz, *Law and Courts' Impact on Democratization* (in Cane, P. and Kurzner, H. (eds) *Oxford Handbook of Empirical Legal Research,* OUP 2010).
[44] Ronald Dworkin, *A Matter of principle* (Harvard University Press 1985).

i) the capacity of an individual to make a decision about medical care and/or treatment; and

ii) how to assess the best interests of someone who lacks capacity to make a decision for themselves about medical care and/or treatment.

Exclusion: cases

i) decided before the entry into force of the MCA in 2007;

ii) concerning capacity to consent to sexual intercourse or contraception;

iii) where the primary focus is an individual's place of residence;

iv) where the focus is on the application of the Deprivation of Liberty Safeguards;

v) with an administrative or procedural focus (such as the allocation of costs);

vi) concerning decisions to which the MCA does not apply (for example, marriage); and

vii) concerning the best interests of a child for the purposes of the Children Act 1989 (assessed by reference to a different test).

I searched two legal databases: Westlaw UK and LexisLibrary (UK). The searches were restricted to cases heard in the English Supreme Court and the CoP, the court with jurisdiction in England and Wales to make decisions under the MCA. The search retrieved a large number of cases (n = 1,596), and the PRISMA flow diagram (Figure 1 below) depicts the process and reasons for exclusion at each stage. The final corpus of cases selected for analysis (n=16) was created by applying reasoned purposive inclusion criteria (see Table 1).

Table 1: Reasoned purposive inclusion criteria for selection of the final corpus of cases

Inclusion criterion	**Reasons**
Decisions of the Supreme Court	The decision-maker of last resort: sets precedent
Cases stated to be of general application	Not restricted to a particular case

Cases where reference is made to palliative care	In case any care/treatment specific findings
Cases where the patient's voice cannot be heard	How to ascertain wishes and feelings and what weight to be given
Cases where specific consideration is given to the significance of the patient's wishes and feelings in the decision-making process	How to ascertain wishes and feelings and what weight to be given

The selection of cases for analysis of judgments was not influenced by the likelihood of the care or treatment considered being offered by a hospice (such cases appeared rarely to come before the CoP), although the provision of palliative care was relevant in several instances. Whilst I had hoped to be able to review cases with a direct correlation to the hospice context, I felt, on reflection, that reviewing a broader range of decision-making places and spaces allowed a more nuanced understanding of judicial development of the MCA, as well as allowing the empirical phases to offer a distinct perspective on the hospice context.

Having identified the final corpus of cases, I took a two-stage approach to analysis. I looked first at the application of the MCA, reviewing how the key legal concepts of capacity and best interests had been interpreted and applied in the judgments (the doctrinal lens). I looked at the rationale for the judge's application of the MCA framework to the facts and recorded my findings in a detailed data extraction table. I then went back to the judgments and read them again, this time approaching them as narratives in which the judge re-told the story of the events leading to the court's involvement (the narrative lens). By combining a doctrinal and a narrative engagement with the content of each judgment, I wanted to develop a deeper understanding of how the MCA was being embedded into practice. Reading through the judgments-as-narratives, I coded each by reference both to the three key themes from the genealogical analysis, (*emergence of the individual*, *a person of value*, and *role of law*) and to note different ideas. This allowed me to identify patterns in the judges' reasoning that reflected

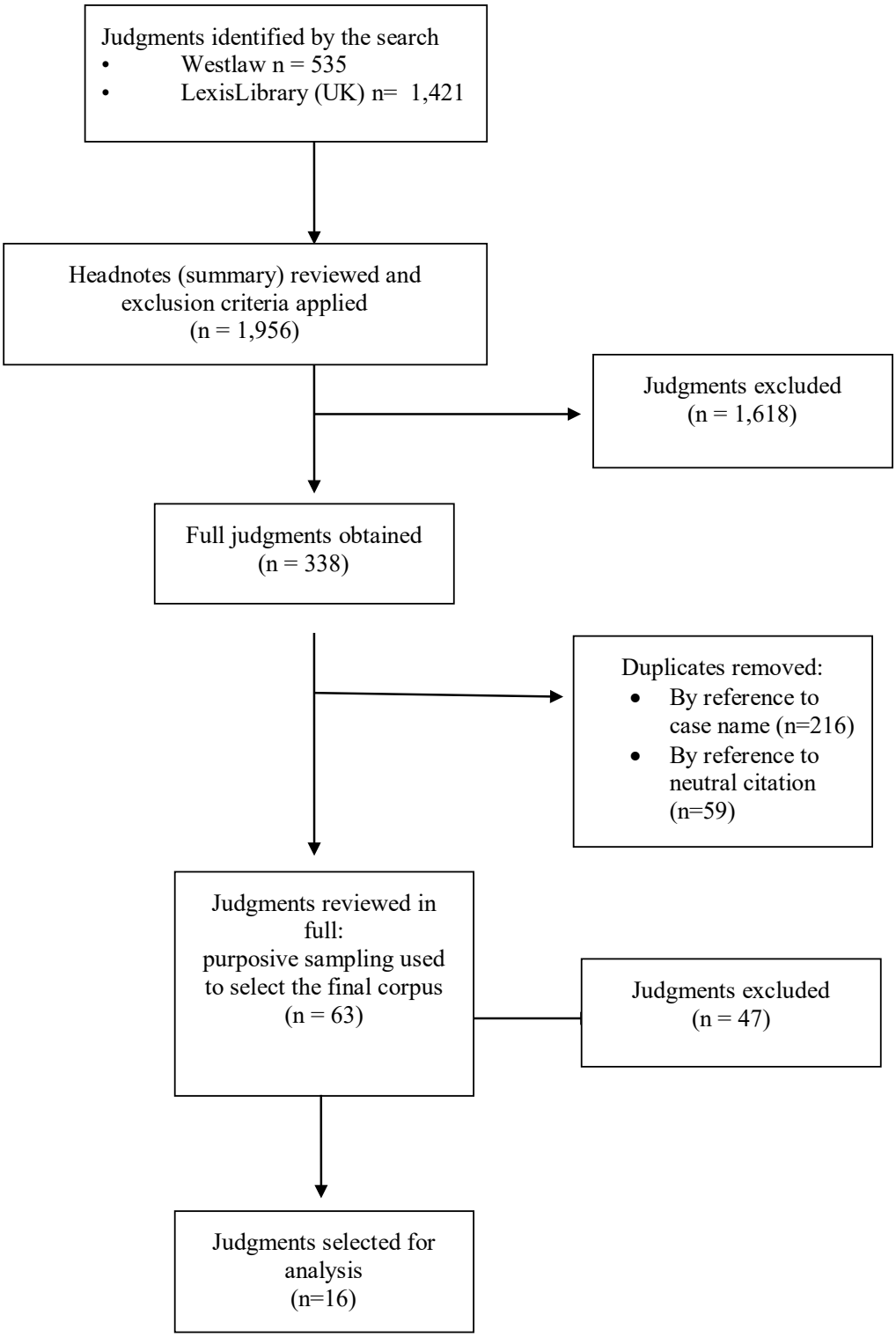
the aims of the legislators, as identified by the genealogical work, and to capture any departures, or differences. The identification of additional patterns as part of the narrative stage facilitated the development of richer theoretical ideas as the study progressed.[45]

Applying this two-stage doctrinal / narrative analytic approach to the final corpus of cases, I found that the three key themes from the archive underpinning the Foucauldian review, *emergence of the individual*, *a person of value*, and *role of law*, continued to be relevant to the way in which the legislation was judicially interpreted in the CoP. The genealogical analysis had captured the themes broadly, and the judgments were the lens which, over time, afforded a closer view of *how* CoP judges considered the person at the centre of the decision-making process (*emergence of the individual*), accorded value to their wishes and feelings (*a person of value*), and brought the MCA to bear in so doing (*role of law*).

Interestingly, the review of judgments revealed a change in judicial engagement with the person whose story was being told (P). Where direct engagement was possible, judges increasingly chose to meet P. This suggested a shift in judicial practice in the CoP, a correlation with the intentions of the legislators (*emergence of the individual, a person of value*) and enabled a theoretical consideration about the constitutive nature of relational engagement in legal practice (*role of law*).[46] This was reflected in the findings of the empirical research in the final phases of the study, where participants' described a consciously relational engagement with patients in their care as an important aspect of their interpretation of the MCA framework.

[45] While a detailed discussion of these findings is without the scope of this article, the point can be illustrated by reference to a theme of time and space in the context of relatives' memories of the former wishes and feelings of someone existing now in a minimally conscious state. In *Briggs v Briggs* (No.2) [2016] EWCOP 53*,* Charles, J. (at [58]) accepted the views of others as evidence of Mr Briggs' wishes and feelings, allowing me to suggest that, in so doing, the judge was acknowledging importance of an idea of *relational* autonomy in MCA decision-making.
[46] For discussion of the theoretical underpinnings see Nedelsky (n 16) particularly Chapter 1.

Figure 1: stages of the case law review



5. Discussions with Hospital Staff

In phase three of my research, I explored how senior staff (decision-makers from across the multi-disciplinary team) worked together to interpret and apply the MCA, finding that the policy and practices they implemented set an important organisational dynamic. Again, I employed a staged approach, reviewing the organisational policy documents for the hospice and then discussing their development and application in practice with a group of senior staff with responsibility for ensuring compliance with the MCA. Two group interviews were held at the participating hospices (see demographic information in Table 2). The group interviews were semi-structured and followed a discussion guide provided to participants in advance. The focus was on the key concepts of capacity and best interests within the broad legal framework of the MCA and the narrower framework of particular hospices' policies and procedures.

Table 2: Group Interviews: Demographic information

	Hospice 1	**Hospice 2**
No. of participants	6	4
Gender: • M • F	 1 5	 2 2
Age range (years)	28-61	36-58
Nature of role: Clinical care Social care Education	 4 1 1	 3 1

In phase three, I was interested in exploring both the bigger picture (the development of the law and the regulatory environment in the hospice sector) and trying to uncover any specific influences on policy and implementation of the MCA in the context of each participating hospice (such as, for example, whether learning from a particular patient's journey, or a staff member's experience was reflected in guidance or training). I thus conceived the organisational 'policy' as being something which was subject to change while remaining part of the ongoing

culture and organisational context of the hospice. I investigated this at an organisational level and, in phase four, in conversation with individual staff members.

Phase four was the final phase of the study. I recruited and interviewed hospice staff from the two participating hospices. Recruitment was by purposive sampling of 3-4 staff members from each participating hospice. I endeavoured to capture a broad range of perspectives, recruiting from across the range of clinical, social care, bereavement and spiritual care professionals represented on the staff, my central inclusion criterion being that all participants had responsibility for supporting patient treatment and/or care decision-making. Inclusion criteria for the sample were:

1. responsibility for assisting patients make care or treatment decisions; and/or
2. responsibility for religious, spiritual (or specifically non-medical) support and decision-making.

Information about the research was publicised within each participating hospice and interested staff members were invited to email me. In the event, only three participants were recruited in each participating hospice (see Table 3 for participant information).[47]

Table 3: Individual Interviews: Demographic information

	Hospice 1	**Hospice 2**
No. of participants	3	3
Gender: • M • F	 1 2	 0 3
Age range (years)	23-53	34-57
Nature of role: Clinical care Clinical care/education Social care	 2 1	 2 1

[47] Recruitment to the study was challenging. Recruitment information was circulated by an organisational contact and Initial expressions of interest were followed up once, in accordance with ethical approval for the study.

In my semi-structured interviews with these staff members, I explored their personal experiences of decision-making under the MCA. Participants were asked to narrate memorable experiences of decision-making for patients, but were not asked to situate their stories in a discussion of the legal or policy context. Rather, I encouraged them to present and discuss their experiences by reference to the principles that guided *their* engagement with care and treatment decision-making for patients. Participants explored the experiences chosen in depth, concentrating on the way those involved (such as the hospice staff, the patient, and their family) took part in the decision-making process. The aim of this final phase was to understand the professional perspective of MCA decision-making, particularly as regards the patient's role in the process. As with Ewick and Silbey's work in investigating legal consciousness, it was in individuals' *perceptions* of the law and not the law itself that I was interested.[48]

Having collected the range of data I hoped would illuminate whether, and, if so, how, the intentions of the legislators had become woven into the present understanding by hospices of the MCA in practice,[49] I drew the threads of the research together in a synthesis of the findings across the study as a whole. In the section that follows, I describe how I used template analysis in each phase of the study, and to synthesise the findings across all four phases.

Using Template Analysis to Analyse and Synthesise

Template analysis, a type of thematic analysis, uses hierarchical coding to develop a coding template, initially based on a subset of the data, which is then revised and refined as it is applied to further data.[50] In each phase of the study, I developed an initial template from my inductive

[48] Ewick and Silbey (n 21). Legal consciousness became a key feature of my research and a significant aspect of the discussion about the distinctiveness of the hospice movement.
[49] For a discussion about historicisng the present, see Penny Tinkler & Carolyn Jackson, 'The past in the present: historicising contemporary debates about gender and education' (2014) 26 Gender and Education 1 70 available at < doi.org/10.1080/09540253.2013.875131> accessed 16 May 2022. See also Series (n 34).
[50] Brooks, J., McCluskey, S., Turley, E. & King, N., 'The Utility of Template Analysis in Qualitative Psychology Research' (2015) 12 Qualitative Research in Psychology 2 202.

coding of the first document, judgment, or interview transcript, which then formed the basis for coding the second, and so on, facilitating the iterative development of more complex codes and, ultimately, themes from the documents, judgments or interviews being analysed. In developing my initial coding template for each phase of the research, I worked through a process, based on that described by Brooks et al.:[51]

1. I familiarised myself with the data by reading and re-reading the documents, judgments or interview transcripts. I preliminarily coded the data by highlighting aspects which seemed to me to be interesting in terms of facilitating my understanding of the data, or to suggest a pattern or trend.
2. Having identified my initial codes, I organised them into clusters around the patterns (themes) they suggested. In phases two to four, during this process of organisation, I considered the relevance of the broad conceptual ideas represented by the key themes from the genealogical analysis (*emergence of the individual*, a *person of value*, and *role of law*).
3. The resulting conceptual themes, arranged hierarchically, comprised the initial template for each phase, which was then iteratively developed as the rest of the data was analysed.

Template analysis was a useful method for this research, facilitating the identification of patterns and common threads over time and across the heterogeneic narrative sources explored. The method also allowed similarities between themes from earlier phases of the research and the later phases to be made visible. This was achieved by using the three key themes from the genealogical analysis, *emergence of the individual*, a *person of value*, and *role of law*, as 'organising themes' for the analysis of the judgments and the interview data. I did not

[51] Ibid.

characterise or use these three themes as 'a priori' themes, in the sense that they represented some pre-determined theory or structure, because to have done so would have created a sense of conflict with the 'bottom up' approach that underpinned my research. However, I reflected on the relationship between these key themes and the data I was analysing at each phase of the project. In this way, analysis of the judgments and the interview data continued the process of 'historicising the present',[52] which my genealogical approach to constructing the life story of the MCA had started. The challenge, from the perspective of the research as a whole, was to bring the analysis together coherently, to make sense of the relationship between the four phases, and to bring the life story of the MCA up to date (or, at least, to the date of the data collection).

I considered a variety of methods to achieve this. Noblit and Hare's approach to synthesis, based on identifying key concepts from one study and 'translating' them into another to take an argument beyond the content of the original studies, appeared the best 'fit'.[53] However, their notion of 'translation' of related concepts into each other felt too static for my purposes, suggesting simply a different means of expression of the same idea. I looked to the work of the relational and narrative theorists, whose ideas had influenced my situation of this research, for a more dynamic approach to synthesis. Barad's ideas about diffraction as an analytic tool were immediately appealing.[54] Initially a quantum physicist, Barad has grounded her concept of diffraction on the behaviour of waves when they overlap, and the apparent bending and spreading of waves when they encounter an obstruction.[55] She uses diffraction as a metaphor for describing an approach to analysis that attends to the effects of difference, suggesting that diffraction makes visible the relational nature of change.[56] Barad's ideas about diffraction

[52] Livholts and Tamboukou (n 24) 64.
[53] Noblit GW and Hare RD, *Meta-Ethnography: Synthesizing qualitative studies* (Sage 1988)
[54] Barad (n 17).
[55] Barad (n 17), see discussion in Chapter 2.
[56] Barad (n 17), see particularly Chapter 2 at 71.

underpinned the way I had visualised the themes from the (phase one) genealogical research moving into the legislative process and, from there, into the MCA itself. Thinking about synthesis, and Noblit and Hare's use of 'translation' to pull related concepts together, I decided to create a 'synthesising template'. I based my synthesis on Barad's notion of diffraction to examine how the themes from the four phases were related (characterising each phase of the research as a slit through which the themes were passing and looking at the diffraction patterns created), and, in so doing, to make visible any (lack of) change in interpretation of the MCA over time.

In order to develop this synthesising template, I reviewed the analytic templates from each of the phases, looking for over-arching links and patterns across the datasets. As I had done in approaching the analysis of each phase, I again considered the relevance of the *emergence of the individual*, a *person of value*, and *role of law* to the data as a whole. This exercise generated four broad themes, within which the themes of each phase nested (see Table 5). The synthesis clearly makes visible the effects of change over time in the interpretation of the MCA, suggesting, for example, that the notions of care, compassion and therapeutic jurisprudence, have become increasingly relevant to law and legal thinking, and making visible the role of legal consciousness in hospice staff's interpretation of the MCA. Interestingly, for the aims of the research (to consider how the MCA's principles have been understood and interpreted in hospice practice, and to compare this understanding to the legislators' intentions), the three key themes from the genealogical analysis, *emergence of the individual*, a *person of value*, and *role of law*, remain visible in the final template, albeit patterned differently after their passage through time and place.

Table 5: The synthesising template: over-arching themes showing links to previous phases of the study

Theme	**Phase 1**	**Phase 2**	**Phase 3**	**Phase 4**
People and the decision-making journey	• Emergence of individual • Person of value • Power	• Emergence of individual • Individual's story • Person of value	• Individual in the decision • Collegiate approach • The hospice movement • People in organisations	• Circles of care • The patient's narrative • Perspective – seeing the world through different eyes • A little bit of me
Place, space and time	• Power over a future narrative	• Me, myself and I • Narrative wormhole • Distance and place	• Hospice/acute/community/home • The patient before	• Hospice/acute/community/home • Patient narrative • Luxury of time • MCA journey – legal consciousness and experience
Law, care and compassion	• Law as a compassionate power • Therapeutic jurisprudence	• **Compassion as integrative theme**	• **Compassion as integrative theme**	• **Nature of care as integrative theme** • Change of care over time (old school/habitus)
Role and nature of law	• Change in focus of law • Rights and responsibilities	• CoP role – socio-legal focus • Therapeutic jurisprudence • Legal consciousness	• Legal consciousness • Law as a resource	• **Legal consciousness as integrative theme**

6. Reflection and Conclusion

The story of my research started with negatives. It started with an assessment that the MCA, a 'visionary piece of legislation for its time', had not been well understood or implemented in practice, that its 'empowering ethos' had not been delivered, that the rights it conferred had not been realised, and that the responsibilities it imposed had not been accepted.[57] The conclusion of my research, however, was positive. My findings suggested that the MCA *is* well understood in hospices, that its implementation *does* empower patients who lack or are losing capacity, as it allows their wishes and feelings to be heard and reflected in decisions concerning their care and treatment at the end of life. Furthermore, I found that hospices (as organisations) and their members of staff were *keenly* aware of their responsibilities to work within the requirements of the MCA.

In this article, I have described how my phased approach to the research enabled the collection and interrogation of varied narrative datasets. My approach was the cumulation of varied readings, discussions, and excursions into methodological dead ends that looked interesting from the entry point. This was both a benefit and a challenge, for obvious reasons. My research design facilitated a coherent approach to a potentially complex combination of methods in a relatively small-scale piece of research. Template analysis worked well in this context, but my use of it was not unusual and need not figure in this reflective conclusion, save to acknowledge its important role in allowing the 'flow' of the themes from one phase to another, and its role in facilitating a synthesis which led to an interesting theoretical discussion.

In terms of the wider application of the methods to socio-legal scholarship, I suggest that there is a place for an adaptation of the systematic review, and for the genealogical approach, underpinned by Foucauldian thinking, that I developed to investigate the 'becoming' of statute law. While the former, in the form I used for my exploration of the phase two data,

[57] Select Committee on the Mental Capacity Act 2005 (n 2).

would not satisfy a systematic reviewer from a social science background, it does allow for a focused and defensible collection and analysis of a body of case law. As to the latter, a genealogical approach has, from a socio-legal perspective, continuing merit for researchers.[58]

In my research, its purpose was to facilitate a detailed exploration of the reasons for which the time was right for the MCA to come about, but it has broader application than this.[59] A Foucauldian genealogical approach could accommodate the collection of a larger documentary archive than I was able to navigate, and, combined with a transparent and replicable search strategy, could defend itself against claims of researcher bias. Such an approach also has the potential to draw out and uncover more complex themes rooted in power dynamics, which were a key fascination of Foucault's.[60] A genealogical approach might be of most value, though, in an investigation that seeks to understand how a *change* in the law came about, as my study did or, as Series' recent work has done, to explore why the law in some areas has changed *less* than we might hope or expect.[61] My current research on the use of direct-to-consumer genetic testing by donor-conceived people,[62] includes a comparative analysis of laws removing gamete donor anonymity in several jurisdictions. I am exploring the ways in which a Foucauldian genealogical analysis of notions related to the importance of genetic heritage in each society prior to, and post, the legislative change might support a comparative socio-legal study. Laws with contested normative framing (assisted dying might be an example) would also offer fertile ground for a genealogical analysis based on Foucault's ideas.

[58] I note Series' use of a similar approach for the research into post-carceral social care, described in her recent book (n 34).
[59] See for example Series' recent book, ibid.
[60] See, for example, Foucault (n 31); Foucault, M. *The Confession of the Flesh* in Gordon, C. (ed) *Power/Knowledge Selected Interviews and Other Writings* (The Harvester Press Limited 1977)

[61] Series (n 34).
[62] See the ConnecteDNA research project website: <https://sites.manchester.ac.uk/connecte-d-n-a/> Accessed 31 August 2022.

Thus, I think that my genealogical approach is generalisable to laws and legislative change both in England and Wales and in other jurisdictions.[63] The same is true of the phased combination of documentary and empirical enquiry, linked and synthesised by template analysis, which I used to investigate the development of the MCA's key concepts over time. My research focused on a particular legal framework in what I have characterised as a unique (decision-making) context. However, my methodological approach would serve equally well as a means of investigating the 'life story' of another statute in a different context.

Looking closer to home for my final reflections, my exploration and development of the research methods I have described above has facilitated a richer engagement with the relationship between law and society, and between people (including judges) and 'the law', particularly in the context of laws that support caring relationships, and specifically where decision-making capacity is in issue.

[63] Interestingly, in Sweden, before the Government can draw up a proposal for legislation, or legislative change, analysis and evaluation of the matter in question is required. The resulting Official Reports (Statens Offentliga Utredningar, or SOU) offer a readily available and comprehensive archive of information to the socio-legal researcher interested in a genealogical enquiry.

Take (what they say) with a pinch of salt: Engaging in Empirical Research to Understand the Parameters of the 'Quality' in 'Poor-Quality Defence Lawyering'

*Ashlee Beazley**

Keywords: Empirical legal research; legal advice; qualitative research

Abstract

The criminal defence lawyer is an essential component to the equality of arms, due process, the enforcement of the prosecution's burden of proof, and to the right to a fair trial. Despite this, little attention has yet to be paid to what the different qualities of legal defence assistance—whether adequate, effective, sufficient, etc.—actually amount to. This article presents the methodology and results from a comparative empirical legal study on the quality of criminal defence advocates in England & Wales and Belgium. The study presented, and the wider PhD research to which it forms a part, seeks to construct a theoretical framework by which poor-quality (insufficient) defence representation may be identified, understood, contextualised, addressed and remedied. To this end, the empirical research undertaken and outlined in the article which follows focuses on a particular source of information: the data acquired from semi-structured interviews held with defence practitioners about their (own) perceptions, opinions and experiences of the quality of defence representation.

The article discusses the extent to which lawyers are a reliable source of data and the usefulness of empirical research as a means by which legal theory may be developed, articulated and tested. If, for example, quality lawyering is to be defined in the hope of demarcating "sufficient" quality from "insufficient", then it is both natural and necessary to involve the subjects of this research, the lawyers. A qualitative empirical study which utilises constructivist grounded theory and critical realism is, this article suggests, one means by which this delineation may be ascertained, one which also seeks to contextualise the data obtained whilst acknowledging the role and effect of the researcher in question.

1. Introduction

As Lord Levenson once astutely noted, it is a 'critical test of the freedom inherent in our democratic society' that those accused of committing criminal offences can and should be represented by 'capable criminal advocates, independent in spirit' who, subject to the rules of law and procedure and to the 'dictates of professional propriety', are prepared to put the interests of their clients at the forefront, irrespective of personal disadvantage.[1] The criminal defence advocate, it can be said, is of fundamental importance to the modern criminal justice system. An obvious statement, perhaps, but this does not detract from the fact they are an essential component to the equality of arms, due process, the enforcement of the prosecution's burden of proof, and the right to a fair trial. The assistance of legal counsel in the preparation of one's defence serves to promote individual dignity—it recognises and upholds the right of the accused to exercise individual participation in their defence,[2] while also protecting the legitimacy and integrity of the proceedings' due process.[3] The presence and assistance of legal counsel, furthermore, help to alleviate the stress of facing charges, aid in ensuring defendants are treated properly during criminal proceedings and, fundamentally, give expression to

* PhD Candidate, KU Leuven. The author would like to thank the two peer reviewers of this article for their useful and insightful comments and suggestions. Any errors or omissions are the author's.

[1] *R. (On the application of Lumsdon & Ors.) v. Legal Services Board & Ors.,* [2014] EWHC 28 (Admin), para. 1, per Lord Levenson. The full quotation is: 'It is a critical test of the freedom inherent in our democratic society that those accused (usually by the State) of committing criminal offences can and should be represented by capable criminal advocates, independent in spirit who, subject to the rules of law and procedure which operate in our courts and to the dictates of professional propriety, are prepared to put the interests of their clients at the forefront and irrespective of personal disadvantage. Similarly, advocates instructed to prosecute crime must be impartial, balanced and fair. These are the values, to the great advantage of the rule of law in this country, that have long been embedded in the practice of advocates before our criminal courts. Those who have the responsibility for the regulation of advocates (whether barristers or solicitors) are imbued with the same sense of the centrality of independence and mindful both of the need to maintain standards and the critical importance of supporting professional independence.'

[2] John D Jackson and Sarah J Summers, *Obstacles to Fairness in Criminal Proceedings: Individual Rights and Institutional Forms* (Hart Publishing 2018) 18.

[3] Herbert L. Packer, 'Two models of the criminal process' *University of Pennsylvania Law Review* 113 (1964) 1, 60. Packer notes that the criminal defence advocate is a "crucial figure" to the Due Process Model of criminal justice, for on their presence depends the viability of the Model's prescriptions (ibid.).

defendants' participation more effectively than defendants themselves may otherwise be able to do so.[4]

Defence counsel not only act as the defendant's voice, they are also tasked with ensuring the most effective case is mounted against the prosecution—their presence is thus also vital in upholding the institutional need for the prosecution case to be challenged.[5] There can be little doubt, then, that the right of the accused to be assisted by a lawyer[6] is of crucial importance to ensuring the legitimacy and integrity of criminal proceedings: the fairness of the criminal justice system, the European Court of Human Rights (ECtHR) has stated, depends on an accused being adequately defended, both at first instance and on appeal.[7] This idea, that an accused should be adequately defended, is reflected in the right and requirement that the legal assistance provided by counsel be effective.[8] Despite this, and notwithstanding the continued focus on increasing the presence and assistance of advocates during all stages of proceedings, little attention has yet to be paid to what the different "virtues" of legal defence assistance—adequate, effective, sufficient—actually amount to.[9]

---

[4] Jackson and Summers, 18.

[5] Sarah Summers, *Fair Trials : The European Criminal Procedural Tradition and the European Court of Human Rights* (Hart Publishing 2007) 110.

[6] The right of access to a lawyer is an essential procedural right of suspect and accused persons in criminal proceedings, as art. 6(3)(c) of the European Convention of Human Rights. The right of access to a lawyer plays a significant role in facilitating other procedural rights, such as the right to the accused not to incriminate themselves, the right to competent and effective legal advice and the right to have adequate facilities for the preparation of a defence. Despite this, there is no "uniform understanding" of the role and function of defence counsel. As Wohlers observes, there is an interesting distinction between counsel required to assist the accused person, and counsel expected to act as the accused's representative; these activities, he argues, are not one and the same: Wolfgang Wohlers, 'The Role of Counsel in Criminal Proceedings, in Jackson and Summers (n 2) 127. For a discussion of this idea, in the context of the right to defend oneself, see Ashlee Beazley, '*Sans l'avocat:* The Right to Represent Oneself and EU Fair Trial Rights' in S. Allegrezza and V. Covolo (eds.), *EU Fair Trial Rights in Criminal Proceedings* (Hart Publishing, 2021, forthcoming).

[7] *Aleksandr Dementyev v. Russia*, App. No. 43095/05 (ECtHR, 28 November 2013), para. 40. In fact, the ECtHR has gone a step further, and held that art. 6 guarantees that proceedings against an accused should not take place without adequate defence representation: *Correia de Matos v. Portugal*, App. No. 54602/12, (ECtHR, 4 April 2018), paras. 122–123.

[8] The right to an effective defence is a part of the right to defend oneself through legal assistance of their choosing under art. 6(3)(c) of the European Convention of Human Rights. See Stefan Trechsel, *Human Rights in Criminal Proceedings* (Oxford University Press 2006) 286–290 and Amal Clooney and Philippa Webb, *The Right to a Fair Trial in International Law* (OUP, 2020) 354–360.

[9] This observation has been by many commentators. Cf., for example, Stefan Trechsel, *Human Rights in Criminal Proceedings* (Oxford University Press 2005) 145. There has also been very little empirical research on the quality of criminal advocacy, despite concerns expressed in jurisdictions like that of England and Wales:

This is surprising: it should be self-evident that the protection of the interests and rights of the accused requires not just the mere presence of a qualified representative, but one who is sufficiently active, vigilant, skilled and competent. If the right to defence oneself through legal assistance is to be effectual, this assistance must be of ample competency to ensure one's rights—not least to a fair trial—are appropriately represented and protected. What distinguishes, in other words, those who provide a suitably effective defence and those who do not? The answer to this, as this article argues, may depend on how these varying descriptions of quality are defined.

The problem with using quality as a qualifying descriptor is that it is a nebulous one—perhaps even an 'essentially contested concept.'[10] Rare is the agreement on how to define it; rarer still is the agreement on how to measure it. Quality is a concept often used and yet it remains ill-determined: it is a word many of us instinctively understand, yet few can precisely explain: how can one define a concept as intangible and variable as that of quality? The *Oxford English Dictionary*, with its explanation of 'a standard or nature of something as measured against other things of a similar kind' suggests a method: that of the comparison. Yet this overlooks a basic, foundational step: in order to have a comparative standard, one must first *determine* the standard.

---

see Gillian Hunter, Jessica Jacobson and Amy Kirby, 'Judicial Perceptions of the Quality of Criminal Advocacy: Report of research commission by the Solicitors Regulation Authority and the Bar Standards Board' (Institute for Criminal Policy Research, Birbeck, University of London, 2018), at < http://eprints.bbk.ac.uk/22949 > accessed 3 June 2022, who note concerns 'about declining standards of advocacy' have been highlighted in reports by the BSB ('Perceptions of Criminal Advocacy', ORC International, 26 March 2012) and the Crown Prosecution Service (CPS) ('Follow-up report of the thematic review of the quality of prosecution advocacy and case presentation', HM Crown Prosecution Service Inspectorate (HPCPSI), March 2012) 1.

[10] First proposed in 1956 by W.B. Gallie, essentially contested concepts are 'concepts the proper use of which inevitably involves endless disputes about their proper uses on the part of their users.' See W.B. Gallie, 'Essentially Contested Concepts' (1956) 56 Proceedings of the Aristotelian Society 167. Applying this idea to the law itself, van der Burg suggests that 'essentially contestable concepts are concepts that refer to ideals or to concepts and phenomena that can only be fully understood in the light of ideals, and that are, as a consequence, open to pervasive contestation.' This may well be the concept of quality. See Wibren van der Burg, 'Law as a Second-Order Essentially Contested Concept' (2017) 8 Jurisprudence 230, 245.

And it is here we reach the crux of the definitional problem of quality: exactly whose standard should be taken as determinative of quality? With personal standards of quality, the answer is easy: that of the respective individual. But with common standards, such as those applied or imposed upon one's defence counsel, the answer is much more difficult. Quality is a useful indicator of the "degree of excellence" demonstrated or possessed by the thing or person in question, so long as it is satisfactorily—and consistently—defined. Its usefulness is rendered futile if conflicts of definition or inconsistencies in application are allowed to remain. Such is the problem of the (presently) inexplicable "quality defence practitioner". While many commentators have sought to define quality,[11] no definition—as it is applicable to criminal defence practitioners—has been agreed upon: the courts and regulatory bodies, for example, each apply different attributes.[12]

The research from whence this article is drawn is a PhD thesis which discuss the quality of defence legal assistance in England & Wales and Belgium[13] and which attempts, at least in part, to proffer a common definition or standard of "quality". A comparative analysis, using England & Wales and Belgium as jurisdictional cases studies, the thesis is concerned with assessing whether the quality of a defence lawyer's performance can be quantitatively

[11] See for example: M Travers, 'Measurement and Reality: Quality Assurance and the Work of a Firm of Criminal Defence Lawyers in Northern England' (1994) 1 International Journal of the Legal Profession 173; Hilary Sommerlad, 'The Implementation of Quality Initiatives and the New Public Management in the Legal Aid Sector in England and Wales: Bureaucratisation, Stratification and Surveillance' (1999) 6 International Journal of the Legal Profession 311; Alan Paterson, 'Peer Review and Quality Assurance Papers Presented at the UCLA/IALS Conference on Enriching Clinical Education' (2006) 13 Clinical Law Review 757; Richard Moorhead, Avrom Sherr and Alan Paterson, 'Judging on Results? Outcome Measures: Quality, Strategy and the Search for Objectivity' (1994) 1 International Journal of the Legal Profession 191; Richard Moorhead, Avrom Sherr and Alan Paterson, 'What Clients Know: Client Perspectives and Legal Competence' (2003) 10 International Journal of the Legal Profession 5.

[12] Although this article will not discuss the standards of lawyering used by the courts and regulatory bodies, it is worth mentioning them in brief for context: in England & Wales, the courts refer to "adequate" defence representation, while the professional codes—the Bar Standards Board and Solicitors Regulation Authority in England; the *Ordre des Barreaux francophones et germanophone* and *Orde van Vlaamse Balies* in Belgium—speak of "competence" and "diligence".

[13] England & Wales was selected as the first jurisdiction as it is both common law and a jurisdiction in which discussion around the quality of defence legal assistance has been ongoing for some time, both in the courts and wider scholarship. Belgium was selected as the comparative jurisdiction as it is a mixed legal system of inquisitorial tradition on which very little has said on this subject. It is also the jurisdiction in which the author is based.

measured, with the objective being to facilitate a defendant's right to access competent and effective legal representation and, in turn, to assist practitioners in more clearly defining the limits of the expectations placed upon them. As a part of the thesis's research, semi-structured interviews with current practitioners on their experiences and perceptions of "quality lawyering" have been conducted. The PhD research is therefore a combined empirical-legal one.

While the broader research looks at a wide variety of source material in pursuit of the aforementioned research objectives, including the jurisprudence of the courts, legal commentary and the regulatory standards of their respective professional bodies, this article will focus on a particular, empirical source of information: the data acquired from semi-structured interviews held with (defence) practitioners about their perceptions, opinions and experiences of the quality of defence representation. First, the empirical research methodology will be presented, followed by a general discussion of (some of) the results acquired so far, before moving on to consider the reliability of lawyers as source of information, particular in the context of qualitative research.

2. Empirical research methodology

The empirical research was initially carried out as an 'exploratory case study',[14] with the intention of testing initial hypotheses regarding the "lexicon" of quality (in the context of defence representation). A case study is an 'empirical study that investigates a contemporary phenomenon within its real-life context', especially when the boundaries between the phenomenon (poor-quality lawyering) and context (wider defence representation) are not

[14] Lisa Webley, 'Qualitative Approaches to Empirical Legal Research' in Peter Cane and Herbert M. Krtizer, *The Oxford Handbook of Empirical Legal Research* (Oxford University Press 2010) 934.

'evidently clear'.[15] It was intended to be exploratory, i.e. a tentative probing of whether defence practitioners are, and could be, a useful source of data regarding the definitions and descriptions that might be given to the different attributes of quality, especially when triangulated against the other sources of data within the research.[16] Accordingly, checking for the availability of, and access to, relevant data (the defence practitioners as sources of data), ascertaining the relevant variables for the research (e.g. the parameters of the "lexicon of quality") and assessing the suitability of the empirical as a study in of itself were additional drivers of the initial research. The empirical research, like that of the wider PhD research of which this forms a part, is comparative in design: it allows improvement of theory building (what is the "quality" in "poor-quality lawyering"?) by comparing the case studies of England & Wales and Belgium.[17] Over time, it became apparent that the defence lawyers were an invaluable source of information and insight, even if what they said—and how they said it—had to be taken with "a pinch of salt."

Between April 2020 and April 2022 16 legal practitioners in England & Wales were interviewed, composed of six barristers, eight solicitors and two practitioners involved in (post-)appellate organisations, one of whom works for the Criminal Cases Review Commission (and who spoke to me in this capacity). In Belgium, 11 practising defence lawyers were spoken to. These interviews sought, through a qualitative research lens,[18] to "capture and categorise social

[15] Ibid., 939.
[16] The use of different data sources collected using a range of research methods assists in 'reducing the possibility that the research will lead to misleading findings based on an incomplete picture'. The process of using multiple data sources to reach 'well rounded conclusions' is known as triangulation: Webley, ibid., 940. The other sources of data within the wider research of the thesis include case law, the professional regulatory bodies' codes of conduct, and secondary scholarship.
[17] Alan Bryman, *Social Research Methods* (5th ed., Oxford University Press 2016) 67. On this point, Bryman goes on to note that the main argument in favour of the 'multiple-case study' is that by comparing two or many cases, a researcher is 'in a better position to establish the circumstances in which a theory will or will not hold.' (ibid.)
[18] The qualitative research methodology is not concerned with whether people or situations are statistically representative. The focus is not on seeking to reach findings that are "generalizable to an entire population", but rather, to utilise the in-depth study to "go beyond description to find meaning." Cf. Webley (n 14) 940.

phenomena and their meanings",[19] namely the views, perceptions and descriptions of the defence lawyer's role, purpose and (quality of) performance, as drawn from the professional experiences of those who were interviewed. While different forms of probability sampling were used to find the respondents, the study itself deployed purposive sampling overall: only criminal defence practitioners were sought in the first instance.[20]

Six barristers were first spoken with, who were contacted via a means of personal connection and referral (known as the "snowballing effect" of participant selection[21]). More barristers were attempted to be reached through criterion sampling[22]: the clerks of those chambers outside of London listed on the Legal500 and similar lists or databases were contacted; of the 51 chambers contacted, none replied. The difficulties faced in finding additional barristers with whom to speak also illustrates the barriers faced when conducting research of this kind: those with personal contacts from which they can benefit will often have greater advantage: less time and effort needs to be put into arranging interviews. Although more success was had when contacting the solicitors (see *infra*), the number of cold contacts

[19] ibid., 928.

[20] Purposive sampling is a strategic, non-probability form of sampling, whereby only those sampled are relevant to the research questions posed. Purposive sampling does not allow the researcher to general their findings to a population. See Bryman (n 17) 408. The deliberate selection of this group of professionals does lend weight to the suggestion the respondents were "elites". In the context of the research presented in the article—that of the data obtained from interviews with defence lawyers—the concept of elite, however, holds less relevance here. The respondents came from a deliberate professional group (defence practitioners) and were chosen specifically because of their membership to this group. Their elite status, in their presumed education and professional prestige, was only relevant to the extent it had granted them membership to their profession. Further discussion on this is unfortunately outside the scope of this article.

[21] A 'technique' by which a researcher initially samples a small group of people relevant to the research question(s) (i.e. practising defence lawyers) and these sample participants thereafter 'propose other participants who have [the same or similar] experience or characteristics relevant to the research'. These participants then suggest others and so on: Bryman (n 17) 411. Perhaps more than other of the other kinds of sampling, snowball sampling has been recognised, as a sampling method, to generate new interactional knowledge: the participants suggested are themselves "recommended" by the suggester, thus given the researcher additional information about these potential participants (the most significant of which may be in relation to the person who suggested them). For more, see Chaim Noy, 'Sampling Knowledge: The Hermeneutics of Snowball Sampling in Qualitative Research' (2008) 11 International Journal of Social Research Methodology 327.

[22] Criterion sample is described as 'sampling all units (cases) or individuals that meet a particular criterion': Bryman (n 17) 409.

made compared to the number of interviews actually arranged and conducted remained significantly disproportionate.

The 11 Belgian defence practitioners were contacted by a means of personal connection and referral as well as through criterion sampling. As the former gave predominantly male lawyers, all of whom work in Brussels and Flanders, the latter sampling technique was deployed to find—where possible—female lawyers, specifically (and particularly) those working in Wallonia.[23] Of the 37 Walloon lawyers contacted (of which approximately half were female), one (male) lawyer responded and was interviewed. The English and Welsh solicitors were sampled using maximum variation sampling.[24] Using online membership lists of organisations like the Criminal Law Solicitors' Association, 355 solicitors were contacted. Of these, 12 responded and of which nine were interviewed.

Each interview was recorded (audio-video) and thereafter manually transcribed.[25] The interviews themselves were semi-structured, with the sample questions being used a catalyst for discussion.[26] The transcript coding was done following the constructivist grounded theory[27] methodological approach to data coding, which suggests grouping interviewee responses by

---

[23] Belgium has two bar associations, the *Orde van Vlaamse balies* (Flemish Bar Association) who represents all Dutch-speaking lawyers in Flanders and Brussels and the *Ordre des Barreaux Francophones et Germanophone de Belgique* (the French- and German-speaking Bar of Belgium) who represents all Francophone (and Germanophone) lawyers in Wallonia and Brussels. To ensure as wide a variation as possible in the experiences, perceptions and background of my Belgian respondents, it was important I attempt to cover both bars (and so, all three regions in Belgium: Flanders, Wallonia and Brussels).

[24] 'Sampling to ensure as wide a variation as possible in terms of the dimension of interest': Bryman (n 17) 409.

[25] For reference, the lawyers and their responses have been anonymised according to the following code:
[First letter of legal role, i.e. **B**arrister, **S**olicitor or general defence **L**awyer]-[Jurisdiction: **UK** or **BE**]-[Geographical location of work: **1** if in capital/main city, **2** if in small city or large town, **3** if practice is primarily rural]-[Order of interview, number]
e.g. B-UK-1-2

[26] As described, in a semi-structured interview, the researcher has a list of question of questions or "fairly specific topics" to be covered, but the interviewee has "a great deal of leeway" in deciding how to reply: Jonathan Grix, *The Foundations of Research* (2nd ed., Palgrave Macmillan 2010) 468. In addition, questions that are not included in the original list 'may be asked as the interviewer picks up on interviewees' replies': ibid. I followed this instinctive approach, and as a result, some questions were asked of respondents that were not asked of others. While all respondents were asked: "what characteristics, for you, make a good/poor-quality defence lawyer?" and "what characteristics, for you, make an (in)effective defence lawyer?", only some were asked "what characteristics, for you, make an (in)competent defence lawyer?" or "what characteristics, for you, make an (in)adequate defence lawyer?".

[27] The purpose of constructivist grounded theory is to learn how people make sense of situations and act on them: Kathy Charmaz, *Constructing Grounded Theory* (2nd edn., SAGE 2014) 9.

building up word-based "codes" or descriptors that are gradually developed into more general concepts and categories which correlate to one's own hypotheses and which relate to the existing literature.[28] The grounded theory method was chosen over analytical methods like classical content analysis or discourse analysis because of its inductive process[29] and emphases on theory building and constant comparison,[30] while constructivist grounded theory, in particular, was selected because it treats research as a construction whilst acknowledging that it occurs under specific conditions of which the researcher may or may not be aware.[31]

While there is always a "conceptual limit" in terms of how one can normatively judge the responses to quality received during interviews,[32] the data obtained from the coded transcripts was interpreted using a critical realism paradigm. Critical realisms 'operates with a different understanding of causation'—it seeks out 'generative mechanisms' that are responsible for 'observed regularities in the social world and how they operated in particular contexts'.[33] Here, in this empirical study, the generative mechanisms of how a defence lawyer performs well (or not well) in their role was sought, in the context of the observed reality of the criminal justice system, its structures, and the presence (as in England & Wales) or absence (as in Belgium) of mechanisms by which to classify defence lawyering as poor. Critical realism is particularly suited to this study because it recognises that human knowledge is fallible, that only a small

[28] Grix (n 26) 581.
[29] Grounded theory involves developing theory as the research proceeds rather than testing a hypothesis posited in advance: Webley (n 14) 944. It begins, in other words, with a broad and conventional understanding of a topic or field and thereafter involves the collection of data from which 'to divine a theoretical account': Sarah Nason, *Reconstructing Judicial Review* (Hart, 2016) 33. Theory is generated and built through the analysis of, and interaction with, the empirical data': Grix (n 26) 113.
[30] As one scholar has noted, the grounded theory method appeals to researchers because it follows the natural pattern of human inquiry: Webley (n 14) 944.
[31] Charmaz (n 27) 13. Charmaz notes that that if we start with the assumption that social reality is 'multiple, processual, and constructed', then we must take into account the researcher's positions, privileges, perspective, and interactions as 'an inherent part of the research reality. It too, is a construction': ibid.
[32] James Thornton, 'The Way in Which Fee Reductions Influence Legal Aid Criminal Defence Lawyer Work: Insights from a Qualitative Study' (2019) 46 Journal of Law and Society 559, 572.
[33] Bryman (n 17) 68. Bryman notes that the multiple case study offers an 'even greater opportunity' with critical realism because the researcher is in a position to examine the 'operation of generative causal mechanisms in contrasting or similar contexts' (ibid.).

part of it may be captured, and that the researcher's active role in the construction of knowledge is an explicit one.[34]

As for the sample size, both overall (27 interviews in total) and by jurisdiction, it is difficult to judge whether saturation has been conclusively reached. Most define theoretical saturation as being when either no new or relevant data is emerging regarding a category of information, or when the 'category is well developed in terms of its properties and dimensions demonstrating variation', and the relationship between the categories is 'well established and validated.'[35] Given, perhaps, the topic of the research (and so, the interviews), to some extent new and relevant information was always being proffered at each interview.[36] That being said, the primary aim of in-depth interviewing (such as that undertaken) is to 'generate data which gives an authentic insight into people's experiences.'[37] As a result, some have argued that the notion of a "sample" is not appropriate here, as every respondent embodies and represents 'meaningful experience–structure links': each is a case study in of themselves, who exist and respond to a particular set of circumstances.[38]

This was the advantage of utilising grounded theory: I was able, after the first few interviews to initially code the transcripts produced and begin analysing the data they contained, thereafter theorising and developing the lawyer's conceptions of quality (and how these related to both my conceptions, and the conceptions offered by the courts and regulatory bodies).[39] In doing so, I was able to identify incomplete understandings and posit possible

---

[34] Anna Pivaty, *Criminal Defence at Police Stations: A Comparative and Empirical Study* (Routledge 2020) 182.
[35] Ibid., 412.
[36] This is also expected where grounded theory (with its emphasis on theory building) is used.
[37] Mira Crouch and Heather McKenzie, 'The logic of small samples in interview-based qualitative research' (2006) 45 Social Science Information 483, 485.
[38] ibid., 493. I could, of course, have continued to interview practitioners in both England & Wales and Belgium. I suggest, however, that the 'implicit requirement' that every case should be taken into account puts the emphasis not on the individuality of each, but 'on the unimportance of the *number* of cases to theoretical explanation' (ibid. Emphasis in original). For my purposes, 27 respondents has proved sufficient.
[39] The grounded theory process following data collection is, to remind: initial coding—focused coding and categorising—theory building—writing up/dissemination. It involves, however, a constant comparative method, meaning theoretical sampling will be deployed to develop theoretical categories: Charmaz (n 27) 18.

explanations for these knowledge deficiencies. Theoretical sampling[40] was then used to fill these gaps while nonetheless progressing towards theoretical saturation[41]: the first interviews identified initial categories of data (concretely: the characteristics and features of each "attribute" of quality enquired about) from which I was able to construct preliminary definitions. This allowed me, in further interviews, to refine those definitions by "testing" various aspects of them, through particular questions or observational (often, comparative) comments to my interviewee(s). With each interview, I explicitly sought to understand how each attribute of quality I asked about was possibly understood by my respondents. Theoretical sampling thus assisted in filling out the properties of each category (attribute of quality), aiding the development of their accompanying (theoretical) definition and, towards the end of the study, demonstrated possible links between the categories, further encouraging comparative analysis.[42]

3. In their words: results of interviews

Without delving too deeply into the results of the interviews themselves, for a full discussion of this is outside the main scope of this article (and demanding of much more space and time), there are a couple of observations I would like to make. The first is the general "definitions" of quality as construed from the interviews themselves.[43] Given the aforementioned

[40] Theoretical sampling means seeking pertinent data to develop the merging theory—its main purpose is to 'elaborate and refine' the categories that constitute the theory: ibid., 193. To be clear, theoretical sampling involves 'starting with data, constructing tentative ideas about the data, and then examining these ideas through further empirical inquiry': ibid., 199.

[41] Theoretical saturation is reached when gathering fresh data no longer sparks new theoretical insights nor reveals new properties of these core theoretical categories: ibid., 213.

[42] Ibid., 205.

[43] In terms of quotations from the transcripts, while the transcripts are an accurate record of the interview recording, in that they include each respondent's own means of speaking (*e.g.* repeating words, using fillers, vocalising agreement during a question), these "quirks" were later edited out, where necessary, to ensure smoother reading of the respective answer, particularly for coding. Text within a square bracket is used to clarify how or what was said, either due to the respondent missing out that connection, or the recording making it difficult to catch.

limitations, I will touch upon two aspects of this: the definition of "good-quality" lawyering and the definition of "poor-quality" lawyering, as ascertained from my interviews (and as naturally dichotomous "properties" of quality).

At the heart of the lawyer's responses to "what is a good-quality defence lawyer?" was a constant: the client, and the lawyer's empathy and communication towards them. This was described as 'it's about can you related to that [your client's circumstances] in some way? Can I give an answer that shows empathy to that person?'[44] or someone who 'listens to what their client wants, explains properly the allegations against them, works out with them how best to progress their case to get them the result which they want.'[45] The client, one barrister observed, 'is a really important human element to criminal law that you need to be aware of as a barrister. And you've got to try and think of the position you client would have been in'.[46] One respondent, in Belgium went as far as to suggest that maintaining client satisfaction is a marker of good-quality: if you obtain the best result possible for the client, they suggested, 'then you deliver good quality and [the client] is happy.'[47] In both England and Belgium, the lawyers saw upholding the client's confidence in them to be an indicator of quality: as one solicitor in England described it—

> 'you need the ability to instil confidence in your clients. You need the ability to be able to see the wood for the trees as I call it; see all the information and know what's important quickly. To be able to get to the crux of what matters and to be able to convey that in a simple way. And to do all that with confidence so that the client trusts that you know what you're doing.'[48]

[44] Transcript, S-1-UK-1.
[45] Transcript, S-2-UK-2.
[46] Transcript, B-1-UK-5.
[47] Transcript, L-2-BE-2.
[48] Transcript, S-3-BE-8.

A poor-quality lawyer, in contrast, while invariably being described as uncaring, was also repeatedly viewed (and so, described) by the lawyers I spoke with as simply "poor" or "lacking": lacking in knowledge, communication skills or simply preparedness. '[Y]ou do [the job] badly if you turn up scruffily dressed without having read the papers, having forgotten part of the papers, [are not] aware of the latest law reports, that sort of thing'[49] or, 'someone who does not turn up to the hearings, is reading boating magazines during prosecutorial proceedings, has done no legal research, gives a superficial defence.[50]' As one solicitor expressed it, the poor lawyer has 'either become jaded or they don't analyse the evidence. They assume that the person sat in front of them must be guilty. They don't read everything and they advise the client wrongly or don't listen to what they've got to say'[51]; or, another succinctly summarised it, poor-quality 'indicates a sort of demonstrable lack of [skills]… to the point where their clients achieve a worse outcome than they [might have] expected if represented by someone who was at least competent.'[52]

The lawyers I spoke with highlighted, clearly, that quality— whether good or poor—is a "slippery concept".[53] Much depends on whose normative frame is used to define it—not only do the courts, regulatory bodies, and defence practitioners all prioritise different values[54] but the lawyers themselves accorded different hierarchical values to different traits and characteristics within their descriptions. While they did, as stated, proffers some grounds of commonality, these were not the only suggestions put forward to the above questions, nor was the client or the lawyer's lack of preparedness (for example) always accorded the same

[49] Transcript, B-1-UK-2.
[50] Transcript, L-BE-1-3. The example of a poor-lawyer 'reading boating magazines' in court is a real one seen by the respondent.
[51] Transcript, S-3-UK-8.
[52] Transcript, S-2-UK-2.
[53] Tamara Goriely, 'Debating the Quality of Legal Services: Differing Models of the Good Lawyer' (1994) 1 International Journal of the Legal Profession 159, 160.
[54] This is a point discussed at length in the thesis itself.

importance by each respondent. Rather, the respondents "on average" mentioned (at least) these aspects of good- and poor-quality defence lawyering, allowing me to accord these features significance (and so, primacy) in my analysis. But it may, indeed, be unfair of me to suggest or infer that these values, and these values alone, are the "most important" markers of good- vs. poor-quality.

For example, also mentioned as "markers" of good-quality lawyering were decision-making and advocacy skills and knowledge of the law. The first was expressed as the need to make 'right strategic choices',[55] or the need to have 'the ability as an advocate', this was also described as 'crucial',[56] a 'first and foremost' quality.[57] The latter (knowledge) was described as 'up-to-date and in-depth knowledge of the law',[58] 'competent in criminal law and procedural criminal law'[59] and knowing 'all the details of the procedure [and] the possibilities of [the case].'[60]

Arrogance, or hubris, was another, specific, trait mentioned by some respondents as being indictive of poor-quality lawyering. This was described, for example, as 'lack [of] judgment [or] inability to form a connection with the jury'[61]; as not bothering to keep oneself up to date with the law 'or assum[ing] that they [the lawyer] know what the right answer is'[62]; 'Complacency. Arrogance. Narcissism. Lack of enthusiasm. Carelessness'[63] or 'lawyers who didn't even know the name of their clients, who didn't know whether they were for the defence or the [prosecution], who did not know anything about the contents of the case.'[64]

---

[55] Transcript, B-1-UK-1.
[56] Transcript, B-1-UK-2.
[57] Transcript, B-1-UK-6.
[58] Transcript, L-1-BE-1.
[59] Transcript, L-1-BE-3.
[60] Transcript, L-1-BE-1.
[61] Transcript, B-1-UK-1.
[62] Transcript, B-1-UK-3.
[63] Transcript, S-1-UK-1.
[64] Transcript, L-2-BE-2.

As an empirical source of data, the lawyers neatly demonstrated Tata's concept of "ethical indeterminacy": the idea that changes in lawyers' practice(s) are 'mediated and negotiated' by a range of competing normative justifications 'about the character of "good" (or even "adequate") defence work'[65] (as seen in their disparate and varied "hierarchies" of "quality values"). 'Ethical indeterminacy', Tata suggests, gives recognition to the highly contestable nature of defence work, in which 'there is a range of very different perspectives defining the idea of "quality."'[66]

Ethical indeterminacy, in other words, suggests that the "proper role" of the lawyer is thus dependent on the normative perspective one takes—a point already made by other scholars.[67] Where Tata develops this idea further, however, is in his proposition that, given there are different ways in which lawyers can approach the same situation in 'a range of reasonably defensible ways', lawyers should be able to draw on differing normative values (and combinations of them) 'as a reservoir of resources of principled justifications for their decisions and advice.'[68] Tata subsequently identifies five broad normative perspectives: that of the traditionalist, 'bureaucratic-efficient', adversarialist, 'radical', and 'client-centred.' For each, the overriding normative value is, respectively: legal knowledge and skill, efficiency, protection from the state, amelioration of 'oppression', and—as the name suggest—client-centred.[69] As they answered my questions, the lawyers I spoke with invariably 'dipped' in and out of each of these normative perspectives. As they did, the answers they gave represent, I suggest, a more holistic picture. More than the courts or regulatory bodies (being the

[65] Cyrus Tata, 'In the Interests of Clients or Commerce? Legal Aid, Supply, Demand, and "Ethical Indeterminacy" in Criminal Defence Work' (2007) 34 Journal of Law and Society 489, 496.
[66] ibid.
[67] See Goriely (n 53) 161–165, who proposes the normative lenses of ""traditionalist," "specialist," "funder," and "consumer."
[68] Tata (n 65) 497.
[69] ibid.

institutions most like to assess, judge or comment upon a lawyer's conduct, performance or quality), the practising lawyers were, I proffer, better able to nuance the conceptions of quality I inquired about. In so doing, they demonstrated to me that the "good-quality" defence lawyer, or the "incompetent" defence lawyer (*etc.)* is not a single thing; they can be many things, some of which may, at times, appear to contradict each other.

An example: some described "poor-quality" lawyering as being both "lacking" (a lack of preparedness, lack of judgment, or lack of knowledge) as well as lazy: 'they've become complacent and they don't care enough so they don't try enough.'[70] Yet as the solicitor quoted expresses—'they don't try enough'—laziness suggests an active motivation on the part of a poor-quality lawyer *not* to prepare or acquire the requisite knowledge. Inherent within this is the suggestion that this lawyer (whoever they are) *could* prepare, for example; only they have chosen not to. This is different to the idea that the lawyer is, on a basic level, actually lacking the ability to prepare adequately, or to acquire the knowledge necessary to perform their role properly. Both "traits" are forms of poor-quality lawyering my respondents suggest. But both are also different *kinds* of poor-quality lawyering.

This brings me to my next observation: it was only through interviewing my respondents that I realised—indeed, learned— that the positive descriptors of quality are not the automatic inverse of their antonym: what amounts to competence, for example, is not necessarily opposite to what amounts to incompetence. Few, if any, of my interviewees, when asked to describe opposing characteristics of quality, did so by automatically giving the inverse or antonym. One barrister, to give but one example, described competence as simply 'not negligent' but incompetence as 'doesn't know the law, doesn't know their own case and I think that does

[70] Transcript, S-3-UK-8.

suggest that they probably don't present it very well either.'[71] This is interesting, for it suggests that normative descriptors of quality (and thus the values on which they are based) may be different depending on whether the quality being described is positive or negative. This is an especially curious point, because while the regulatory bodies, for example, tend to describe their expectations of the lawyers they regulate along positive lines ("competent", "diligent"), the courts, in contrast, focus on the negative descriptors ("inadequate defence", "flagrant incompetence"). Quality, then, cannot necessarily be defined by what it is not, nor can poor-quality be defined by simply stating it is "not quality". This point is, of course, little more than an untested hypothesis at present. But it has opened a possible line of inquiry I may not have found, had I not engaged in empirical research.

Likewise (eventually) requiring further elaboration, but nevertheless worth mentioning briefly here is the discussion around what I term the "factors of influence" on the behaviour (and performance) of defence practitioners. I directly asked each of the practitioners what, to them, might affect or influence whether a lawyer is able to perform their job on any given day (or at any given moment). For those in England & Wales, such factors include the impact of (decreasing) Legal Aid funding (which many scholars now observe correlates to a 'lack of morale' and 'alienation' for defence practitioners[72]), and the effect of the Criminal Procedure Rules, first introduced in 2005, which have—arguably—strengthened the lawyer's duty to the

[71] Transcript, B-1-UK-3.

[72] Cf. Thornton (n 32); James Thornton, 'Is Publicly Funded Criminal Defence Sustainable? Legal Aid Cuts, Morale, Retention and Recruitment in the English Criminal Law Professions' [2020] Legal Studies <https://doi.org/10.1017/lst.2019.31>; Lucy Welsh, 'The Effects of Changes to Legal Aid on Lawyers' Professional Identity and Behaviour in Summary Criminal Cases: A Case Study' (2017) 44 Journal of Law and Society 559; Daniel Newman and Lucy Welsh, 'The Practices of Modern Criminal Defence Lawyers: Alienation and Its Implications for Access to Justice' (2019) 48 Common Law World Review 64; Tom Smith, '"Justice For Sale": An Empirical Examination of the Attitudes of Criminal Defence Lawyers Towards Legal Aid Reform' <https://pearl.plymouth.ac.uk/handle/10026.1/9003> accessed 3 June 2022. Both Thornton, and Newman and Welsh observed an increasing dislike amongst defence practitioners over the ways in which decreased funding was restricting their ability to perform their role to the best of their abilities. Thornton writes that 'what seemed to make [the lawyers' unhappy was the way [the] lower fees forced them to work) to keep their practices solvent—often in ways that they found restrictive, distasteful or even unethically uncomfortable' (Thornton (2020) 22).

court at the expense of their duties to the client.[73] For the lawyers in Belgium, such factors include training (one respondent described insufficient training as being akin to a young trainee, 'even if clever and bright', becoming 'like a young sapling that grows without the support of a stake'[74]) and one's relationship with their client, for whom, ultimately, the defence lawyer is representing and acting.[75] While further analysis on this is—for the moment—beyond the scope of this article, a final observation should be made that "quality", however defined, cannot be considered in isolation. The context surrounding a legal practitioner can be extraordinarily influential, whether they are much aware of this or not. An obvious remark, perhaps, it was nonetheless my respondents who directed me to—or at least, emphasised, through their answers—this point.

4. A pinch of salt? Reliability of lawyers as sources

In an article about empirical legal methodology, it is pertinent, I think, to discuss both the virtues and possible pitfalls of the methodology under examination. And there are a number of drawbacks to be brought to attention. First, the limitations of the "interview" as a form of data gathering: as an "obtrusive" method, it elicits the statements to be analysed—nothing a respondent says can be viewed in "contextual isolation" without also accounting for the influence and interaction of the interviewer.[76] Interviewing is also "individually-focused",

[73] For further discussion on this, see Mike McConville and Luke Marsh, 'Adversarialism Goes West: Case Management in Criminal Courts' (2015) 19 The International Journal of Evidence & Proof 172; Tom Smith, 'The "Quiet Revolution" in Criminal Defence: How the Zealous Advocate Slipped into the Shadow' (2013) 20 International Journal of the Legal Profession 111; Ed Cape, 'The Rise (and Fall?) Of a Criminal Defence Profession' [2004] Criminal Law Review 401; Ed Cape, 'Rebalancing the Criminal Justice Process: Ethical Challenges for Criminal Defence Lawyers' (2006) 9 Legal Ethics 56; and Ed Johnston, 'The Adversarial Defence Lawyer: Myths, Disclosure and Efficiency—A Contemporary Analysis of the Role in the Era of the Criminal Procedure Rules' (2020) 24 The International Journal of Evidence & Proof 35.

[74] Transcript, L-BE-1-3.

[75] Transcript, L-BE-2-2.

[76] Crouch and McKenzie (n 37) 486. Crouch and McKenzie note that it is for this reason 'that the literature on interviewing contains extensive discussions of problems associated with the interactive aspect of the interviewing process.' (ibid.)

which can lead one down the 'slippery slope of methodological individualism' when it comes to explanations.[77] It can be easy to attribute the rationale or explanans to an individual, rather than to a field-level or relational explanation.[78] In the context of this research, for example, I must ask whether the insights I have gained from the practitioners with whom I spoke are to be credited or attributed only to the practitioners (as—self-acknowledged—individual examples of "good-quality" lawyers) without providing a wider system or contextual analysis. I cannot, for example, easily identify or attribute the extent to which their responses are influenced by (and so, came about because of) their training, professional experiences and the general vernacular which surrounds their work, or whether their perceptions and views on the topic stem from inherent values and traits unrelated to their profession. Differences between the foci of the English and Welsh lawyers' answers compared to those given by the lawyers in Belgium were also evident: the Belgian lawyers generally characterised good-quality lawyering, for example, as meeting the client's expectations, as opposed to acting in their client's best interests (a common suggestion in England and Wales). Further researcher is needed to understand the potential reasons behind these differences: are they due to differences in education, training and legal culture, for example, or because Belgium more readily allows civil proceedings to be brought against a lawyer by dissatisfied clients?

Furthermore, my role, as the interviewer and researcher, must be recognised. I see insights in my respondents' answers, for example, in part because they use the vocabulary that I both expect and am myself, as a legal research, familiar with (and because my own experiences and training have given me this familiarity). This too is a limitation of interviews: they 'encourage

[77] Michèle Lamont and Ann Swidler, 'Methodological Pluralism and the Possibilities and Limits of Interviewing' (2014) 37 *Qualitative Sociology* 153, 162.
[78] ibid.

us to find coherence in narratives and worldviews.'[79] In other words: I see answers to my questions in part because I infer that they are there and that they can be used. Qualitative research like interviewing relies on an inductive form of reasoning, deriving general themes or patterns from the data collected.[80] As the interpretivist school of thought argues, one's analysis will always reflect one's own frame of reference; as such, it cannot be disregarded that 'socially significant data' do not (in fact) 'exist independently of the researcher'[81] and that interview material is, therefore, 'ultimately comprehended within a frame of situation', a social milieu which is assumed to exist ontologically prior to both the respondents' and the interviewers' actions 'and therefore causally related to them.'[82] I must acknowledge, therefore, that a different researcher, asking the same questions, could both elicit different responses *and* draw different conclusions. Nonetheless, whilst acknowledging this limitation, I chose to utilise both constructivist grounded theory and the critical realism paradigm in the hope I might both understand and attempt to explain the phenomena I am researching—the quality of defence lawyering—while still paying heed to the specific role my presence, as the researcher, has played in the obtaining and interpreting of the data acquired.[83]

This is where the title of this article comes from: that the data and analysis produced from empirical research should be taken with a "pinch of salt": for just as it has been influenced, however accidentally (or subtly) by the researcher themselves, so too is dependent on the context in which the responses are given and recorded (of which I too am part). An interviewee who consents to being both anonymised and recorded is aware that they are being watched,

---

[79] ibid.
[80] Webley (n 18) 929.
[81] Reza Banakar, 'Reflections on the Methodological Issues of the Sociology of Law' (2000) 27 *Journal of Law and Society* 273, 275. The institutional context of subject-matter, including professionalism affect, Banakar notes, 'how studies are actually conducted and the form of knowledge which is produced.' (ibid.)
[82] Crouch and McKenzie (n 37) 485–486.
[83] While strict interpretivists do not strive to establish causal explanations (their emphasis is on understanding), critical realists, in contrast, generally seek to both understand and explain the social world: Grix (n 26) 83, 86.

noted and recorded; this can also lead to "selective recall" or "performative engagement". On the former, I cannot overlook the possibility that my respondents, in answering my questions about quality, selectively recalled experiences which presented them as (always) of good- or admirable quality, and never of poor- or questionable quality. This remark stems, in part, from the observation that *none* of the 27 lawyers I spoke with voluntarily admitting to having ever been (even, perhaps, excusably or understandably) a "poor lawyer" themselves, yet all noted they had "seen" poor-lawyering. By the same token, each of the lawyers I spoke with was just that—a (trained) lawyer. Lawyers are known to engage performatively in their professional work, for (especially, in defence) their role requires them to present as both partisan and neutral as the situation allows.[84] I cannot, therefore, overlook the fact that in agreeing to be interviewed, and in responding to the questions I posed, my respondents engaged in a similar kind of "performance", with me. This was a known limitation to this research methodology: in asking the lawyers to identify what quality is, to them, I risked that their responses would reflect their own values and practices, with little (or admitted) self-awareness as to whether they were, or had even been, of poor quality themselves.[85] As one scholar wryly observed, writing on the incompetence of counsel as a ground for appeal, applications for appeal drafted by counsel 'very rarely allege error by the applicant's lawyers in marked contrast to applications drafted by convicted persons.'[86] The lawyers, it is suggested, are not quick to find fault with themselves.

---

[84] On this idea, see Smith (n 73).
[85] This observation has been observed in both socio-legal and legal studies on the topic. See, for example, Mike McConville, Jacqui Hodgson, Lee Bridges and Anita Pavlovic, *Standing Accused: The Organisation and Practices of Criminal Defence Lawyers in Britain* (Clarendon Press 1994), in particular Chapter 3 on 'The Culture of Criminal Defence'. See also the studies referenced at fn. 11.
[86] Rosemary Pattenden, *English Criminal Appeals, 1844–1994: Appeals against Conviction and Sentence in England and Wales* (Clarendon Press 1996) 105–106.

5. Conclusion

Many view legal decision-making as an 'esoteric activity primarily concerned with the judgement of values' as referenced against or with legal valid 'prescriptive positions and standards.'[87] This activity, so it goes, cannot be scrutinised by empirical investigations.[88] As judgment or assessment of a defence lawyer's quality naturally includes judgment or assessment of their decision-making, does this too mean that this, their quality, cannot be empirically inspected or measured? A lawyer's "quality", it is said, is something more nuanced, something more complex, than merely a 'a standard or nature of something as measured against other things of a similar kind'.

And yet, if we consider the legal practitioner's understanding of the law as not being predicated on 'substantive law alone', but as also being based on their first-hand experience of legal practice, then we also see those that apply and uphold the law are speakers of not a single, concrete body of legal rules, a single 'language game', but multiple language games, and so, multiple concepts.[89] Put simply: the legal practitioner does not "just" apply the law, they apply the law knowing how the law is used in the 'day-to-day life of the judicial system': as they advise their client, or prepare their case, the (defence) lawyer engages in a prediction of how the other legal actors involved (the prosecution, the police, the judge) will also behave.[90] And it is, in fact, to this almost "sociological" assessment that (my) empirical research seeks to get at. *How* a lawyer reaches a decision (particularly on their defence strategy) can reveal much about what they prioritise in the case before them. To repeat Tata's point, made earlier in this article, as there are different ways in which lawyers can approach the same situation, they

[87] Banakar (n 81) 281.
[88] ibid.
[89] ibid., 282–283.
[90] ibid.

should be able to draw on differing normative values (and combinations of them) to justify and explain their advice and decisions. And yet, we also need to be clear on which values are justifiable and which are not. Which, in other words, suggest a lawyer may be of sufficient "quality" (and so made the "right" decision(s), even if the outcome was not to their client's, or the court's liking) and which may not (and who, therefore, made the "wrong" decision(s)). In order to do this, and in order to determine, therefore, whether "quality" is in fact the right parameter or metric by which to make this assessment, one needs to hear from, and so learn from, those making the decisions in question: the legal practitioners themselves.

The natural limits of empirical research methodologies must also be mentioned: empirical research proceeds on the assumption that the data gathered constitutes 'all that is objectively true' about the topic in question.[91] This is, of course, a false assumption, for there may be truths about the topic—here, poor lawyering—that cannot be discovered or refuted by empirical data alone.[92] This is not to undermine the value of empirical research, however. Rather, however imperfectly, this article has sought to demonstrate the usefulness of empirical research as a means by legal theory may be developed, articulated and tested. While theories founded on empirical evidence might be seen as parochial and dependent on context, the law is, first and foremost, a social construction.[93] To this end, its theories cannot be advanced in isolation from its interpretation and application: rather, such theories can—and should—anchor themselves in the social context of the law, its day-to-day practice and experience, as ascertained by empirical research methodologies. This is the value of empirical data. And this, I suggest, is the value of this study: if we are to define quality lawyering in the hope of demarcating

[91] Nason (n 29) 33.
[92] Ibid.
[93] Ibid. See too Galligan, who suggests that while 'empirical studies certainly support the idea that, while rules guide and structure the legal environment, their application involves other social factors, elsewhere described as contextual contingencies, entering into and forming part of that environment': DJ Galligan, *Law in Modern Society* (Oxford University Press 2007) 47.

"sufficient" quality from "insufficient", then it is both natural and necessary to involve the subjects of this research, the lawyers. A qualitative empirical study which utilises constructivist grounded theory and critical realism is, this article suggests, one means by which this delineation may be ascertained, one which seeks to contextualise the data obtained whilst acknowledging the role and effect of the researcher in question.

As one practitioner I spoke with observed—

> 'the problem is that there is a lot of poor-quality representation and there doesn't really seem to be a way to kind of stop that happening.'[94]

This study, and the wider PhD research to which it forms a part, is an attempt to do just this: it strives to construct a theoretical framework by which poor-quality defence representation may be identified, understood, contextualised, addressed and remedied. To this end, the empirical research undertaken and outlined here in part, forms a key component: one might even go as far as to say the lawyers' contributions are worth their weight in salt.

[94] Transcript B-UK-1-6.