Observing Interoperability: A Behavioural Framework and Analysis of Multi-Agency Interactions in the UK Emergency Services

[ACCEPTED MANUSCRIPT - TO BE PUBLISHED IN *SAFETY SCIENCE*]

Richard Philpot[1]

Mark Levine[1]

Charlotte Betts[2]

Megan Nelson[1]

Eloise Duck[3]

Nicola Power[2]

Author Note

Corresponding author: Dr Richard Philpot, Psychology Department, Fylde College, Lancaster University, Lancaster, LA1 4YF, UK. Email: r.philpot@lancaster.ac.uk

[1] Department of Psychology, Lancaster University, UK

[2] Department of Psychology, University of Liverpool, UK

[3] The School of Computing, Newcastle University, UK

This research was supported by grants from the Centre for Research and Evidence on Security Threats (CREST)

Acknowledgements: We thank the delegates of the JESIP Research Symposium, which brought together delegates from academia, policy and practice to discuss challenges to interoperability, for their discussions and feedback around the findings of this work. Practitioner feedback from this event included interest in adapting the behavioural framework for structured observation during live training exercises; the potential to use the findings on physical clustering and siloed communication as reflective prompts during post-exercise debriefs; and the importance of sustained academic-practitioner collaboration to support evidence-informed improvements in interoperability training and evaluation. A read-out from the event can be found here: https://crestresearch.ac.uk/resources/jesip-research-symposium-summary/.

Abstract

In the context of UK Emergency Management, interoperability among the emergency services is essential for effective response to disasters. This study employs qualitative and quantitative observational methods, to introduce a behavioural codebook that measures physical and psychological manifestations of interoperability in cross-services operational meetings, involving the police, fire and ambulance services. Applying this coding framework, we measure the frequency of interoperability behaviour in team meetings—including markers of trust, identification, goals, communication, and flexibility (Power, Alcock et al., 2024). Through a series of multiple logistic regressions, we determine the impact that interoperability behaviour has on subsequent team interactions. Analysis of 5 high quality videos and 102 video stills showed a notable tendency for physical clustering within service units. Effective communication was the most frequently observed component of interoperability, yet it also reinforced intra-group siloed interactions. Goals were frequently discussed, and these mentions were associated with a decreased likelihood of individuals paying attention to, speaking to, and being openly receptive to members of their own service unit over others. Conversely, markers of shared identity were associated with increased cross-service physical clustering—though this did not significantly increase cross-service verbal communication or open gesturing. These results underscore the complexity of achieving genuine interoperability and the need for targeted strategies that address both operational and psychological barriers. Our study contributes to the development of practical measures for assessing and enhancing multi-agency interoperability, essential for improving emergency response coordination.

Keywords: Interoperability, emergency services, video observation, multi-team systems, teamwork

Introduction

During major emergencies and disasters, it is essential that the teams and organisations tasked with responding can combine their expertise and work together as an efficient joined-up team. Recent high profile disasters in the UK – such as the Grenfell Tower Fire (2017), Manchester Arena Bombing (2017) and London Bridge Attack (2019) – have sadly illustrated the adverse impact of fragmented emergency response whilst highlighting the critical need of effective multi-agency coordination. Indeed, following the Manchester arena attack, where 22 victims lost their lives, a public inquiry concluded that failures of joined up teamwork contributed to loss of life on the night (Saunders, 2022). In the UK Emergency Management context, this joined up, multi-agency teamwork has been termed interoperability, defined as "a shared system of technology and teamwork built upon trust, identification, goals, communication and flexibility" (Power, Alcock, Philpot & Levine, 2024, p. 233). Interoperability supports teamwork by ensuring team members have a clear understanding of the structural components and needs of the multi-team system, while also fostering psychological and social connections to support collective sensemaking, problem-solving and decision-making. Effective interoperability between multi-agency teams is a core feature of emergency management plans across the world. For example, the Australian Emergency Management Arrangements Handbook seeks to build interoperability through use of a common language and coordinated national principles (Australian Institute for Disaster Resilience, 2023), and following the diverse and uncoordinated responses across nations to the COVID-19 pandemic, the need for coordinated global cooperation around public health has been highlighted (Khorram-Manesh, et al., 2023). In the UK, interoperability has been identified as a central feature of disaster resilience (Power, Alcock et al., 2024), as discussed in the UK Government Resilience Framework (Cabinet Office,

2022), Strategic Policing Requirement (Cabinet 2023a) and the latest CONTEST counter-terrorism strategy (Cabinent Office, 2023b), amongst others. However, understanding what interoperability looks like and how it can be measured remains a challenge. The dynamic and often unpredictable nature of emergency situations adds further complexity to understanding and quantifying interoperability.

To address this, we introduce a novel behavioural coding framework designed to capture real-time observational markers of interoperability during live multi-agency meetings. Unlike prior work relying on self-report or anecdotal observation, our structured approach allows for a systematic and fine-grained analysis of how interoperability may be behaviourally enacted and how it shapes subsequent inter-agency engagement. What makes this approach distinctive is its integration of video-based observation with theoretically grounded psychological constructs, which enables a more objective and replicable means of assessing how core dimensions of interoperability behaviourally manifest in practice. Building on this framework, the current research explores how multi-agency training colleagues behave during group meetings, via analysis of verbal and non-verbal behaviours during a large-scale training exercise. This analysis will inform the development of a behavioural codebook (Hillen et al., 2024), aimed at identifying markers of interoperability in such meetings. Utilising a combination of qualitative and quantitative observational methods, we systematically analysed interactions during cross-service operational meetings in large-scale multi-agency training exercises. The objective was to assess whether physical and verbal markers of interoperability could be measured in these settings, and further whether these markers could lead to improved cross-service interaction post-meetings.

Our research questions were:

1) How do team members physically interact with one another following multi-agency meetings?
2) Can we observe and code the structural and psychological components of interoperability (trust, identification, goals, communication, flexibility) in these settings?
3) How do interoperability behaviours influence subsequent interactions among emergency service personnel?

Understanding these dynamics is key to developing strategies that effectively counter the natural tendency towards ingroup self-segregation and for promoting genuine interdisciplinary collaboration.

The current study addresses a critical gap in the interoperability literature – specifically, the lack of systematic, real-time behavioural methods for capturing how interoperability is enacted during multi-agency emergency service meetings. Whereas most existing research relies on retrospective self-report accounts or unstructured observation, we introduce and apply a novel behavioural coding framework that captures both verbal and non-verbal indicators of five components of interoperability. By analysing real-world video data from large-scale training exercises, the current study offers both methodological innovation and practical insights for improving interoperability training. The findings have direct implications for emergency response policy and practice, particularly in identifying interaction patterns, such as self-segregation and siloed communication, that may undermine joint working.

One significant challenge to interoperability is fostering effective interaction among diverse teams, as shared goals do not always translate into integrated action (Power & Alison, 2017). Work in psychology suggests a natural tendency towards self-segregation,

even in neutral settings (e.g., Clack, Dixon, & Tredoux, 2005; Dixon & Durrheim, 2003; Schrieff et al., 2005). The current study investigates whether such tendencies extend to professional environments, particularly in emergency service interactions. Specifically, we will analyse how individuals from the police, fire and ambulance services interact, focusing initially on physical proximity patterns. The objective is to determine if there's a significant tendency for personnel to cluster with their own service members beyond random chance. This is significant, as prioritising one's own service members may lead to self-segregation based on professional subgroups, rather than promoting a unified emergency response team. Furthermore, silo working has been identified as a key cause to teamwork failures during emergencies, as demonstrated in the Manchester Arena Inquiry (Saunders, 2022). Determining the tendency for clustering will provide insights into inherent biases in team dynamics and inform strategies for promoting more integrated emergency response efforts.

Additionally, we aim to assess the measurability of behavioural markers of psychological interoperability components (trust, identification, goals, communication, and flexibility) as proposed by Power et al., (2024). We will explore the extent to which the enactment of these components in multi-agency meetings influence interactions across different emergency services. By observing interactions during multi-agency meetings, we intend to evaluate if these interoperability components are observable and if they can bridge gaps between groups. This research will not only validate our behavioural codebook but also provide a more objective and standardised way of measuring interoperability.

Literature review

In the United Kingdom, the emergency services function within a complex environment, necessitating a significant level of interoperability, particularly in response to major events (Pollock, 2013). This interoperability goes beyond mere cooperation,

representing an operational convergence underpinned by human and technological aspects of trust (e.g., Roud & Gausdal, 2019), identification (e.g., Davidson et al., 2023a), goals (e.g., Locke & Latham, 2006), communication (e.g., Waring et al., 2020) and flexibility to adapt (e.g., Curnin & Owen, 2014) (Power, Alcock et al., 2024). It has further been found at strategic decision-making levels that psychological dimensions including psychological safety, critical thinking, expectation management and adaptation are key to crisis management (Curnin et al., 2023). Such a framework is pivotal for ensuring seamless collaboration and a swift response among diverse organisations during emergencies. In response to these challenges, the Joint Emergency Services Interoperability Programme (JESIP, 2013, 2016, 2021) was established. Its goal transcends improving collaboration - it aims to fundamentally address and rectify previous shortcomings in joint efforts. Specifically, JESIP aims to create a more integrated approach to teamwork, seeking to ensure that emergency services can respond more cohesively in future crises. This proactive approach to build better teams aligns with recent calls to action stressing how teamwork doesn't simply happen but must be explicitly taught and prioritised to improve organisational collaboration (Grote & Kozlowski, 2023).

The existing literature extensively discusses the challenges inherent in achieving interoperability among UK emergency services, primarily due to varied organisational cultures, command structures and operational procedures. These differences can lead to inefficiencies in coordinated responses during critical incidents. Allen, Karanasios, and Norman (2014) analysed secondary data on the operational practices of multiagency emergency services to examine how sub-organisational norms impact inter-agency information exchange. They found that despite sharing the common goal of incident management, each service tended to operate independently, influenced by their own

cultural-historical context (see also Charman, 2014). For instance, police focus on routine security and law enforcement, while ambulance services prioritise medical aid. These cultural disparities disrupted communication and were subsequently identified as the most significant obstacles to successful interoperability and information exchange. Work by Power and Alison (2017) further found that despite all emergency teams perceiving a shared "save life" goal, the translation of this goal into intra-agency objectives led to contradictory behaviours (e.g., police wanting a controlled zone to neutralise threats, whilst paramedics want to reach patients).

Moreover, the literature highlights the significant role of systemic, organisational and interpersonal barriers in hampering effective interoperability. Power, Philpot et al., (2025) discuss this with reference to a 'principle-implementation' gap. Their interviews with expert commanders across the emergency services show that while individuals are keen to embrace interoperability, these efforts are restricted by barriers at multiple levels. Specifically, a dearth of funding and the limited scope of JESIP were identified as systemic issues. Problems of incompatible organisational structures represented an organisational challenge, while experienced stress and work pressure disrupted relationships at the interpersonal level. Addressing psychological barriers, such as building trust, effective communication and shared team identification between services have been identified as potential solutions to mitigate these adverse effects (Power, Alcock, et al., 2024). Kapucu (2006) examined the inter-agency communication between public, private and nonprofit organisations following the September 11th World Trade Center disaster. This study determined that pre-established inter-agency relationships and open communication channels led to more efficient crisis management and improved resource allocation. Davidson et al. (2023a) studied the role of shared social identity in enhancing

interoperability among UK emergency services during the COVID-19 pandemic. Their analysis of 52 semi-structured interviews with strategic and tactical responders showed that a sense of 'common fate' cultivated a strong shared identity across various emergency service groups. This shared identity, in turn, improved multiagency collaboration by overcoming traditional interoperability challenges, such as communication and coordination issues.

Despite a common focus on public safety, achieving authentic integration among diverse groups remains a complex challenge (Power, Alcock et al., 2024). Insights from the literature, reveal difficulties in uniting distinct groups with specialised expertise. Sanders (2014) observed that emergency services often operate based on their specific expertise, rather than collaborating as a unified team. This division is made further salient at a basic visual level, where different uniforms worn by emergency responders make each service easily recognisable, but in turn reinforces distinct roles, which may limit unity across the varying services (Curnin et al., 2014). Mitchell et al. (2011) underscored the need for a careful balance between subgroup identities and overarching superordinate identities to maximise the efficiency of multi-team systems. They suggested that an excessive focus on subgroup identities could lead to siloed work practices, whereas too much emphasis on superordinate identities may create identity threats and uncertain professional boundaries. They, along with subsequent researchers (e.g., Power, Alcock, et al., 2024), emphasise the importance of fostering a team identity that integrates both intra-agency aspects (i.e., what defines us as members of a specific emergency service) and inter-agency aspects (what defines us as part of the broader emergency services community). This approach helps in creating a cohesive identity that encompasses both the distinct and collective roles within emergency response teams.

One important way in which the emergency services can improve their multi-agency working is through inter-agency training. Inter-agency training programs can familiarise personnel with one another's roles, responsibilities and operational procedures (House, Power & Alison, 2014). These trainings foster a collective sense of ‘teamness' (see West & Lyubovnikova, 2012) between the different emergency service groups and can bolster interpersonal relationships, which may facilitate more effective collaboration during future emergencies (Power, Betts, Philpot & Levine, 2024). Work by Kirkham (2009) further emphasises the importance of repeated multi-agency training exercises, showing that frequent training significantly enhanced trust across services, and in turn the speed and effectiveness of decision-making. Furthermore, these training events provide advantages beyond the immediate attendees – they offer an invaluable platform for researchers and evaluators to observe and analyse the dynamics of inter-agency engagement in real-world scenarios and the nuances of how different emergency services collaborate, interact and operate in unison. It is this unique observational opportunity that the current study leverages to gain deeper insights into effective strategies for enhancing interoperability among emergency responders.

Although the integration of psychological components and social identities into inter-agency training offers significant benefits (Power, Betts et al., 2024), their inclusion remains uncommon. This omission is significant, as a focus on the subgroup can lead to self-segregation based on professional subgroups rather than fostering a cohesive emergency response team. This tendency towards professional self-segregation mirrors findings in broader social psychology research. For instance, observational research on physical gatherings of individuals in public reveals that even in seemingly neutral settings, like cafeterias and beaches, individuals and groups tend to ‘self-segregate’, gravitating towards

those with perceived similarities and distancing themselves from those who are different (Durrheim & Dixon, 2005; Schrieff et al., 2005). This natural propensity for self-segregation may potentially extend beyond casual environments into professional settings, including emergency service meetings. Specifically, emergency services teams, despite overarching common goals, may unconsciously flock towards their own subgroups, creating invisible barriers to sustained interoperability.

Beyond physical proximity, individuals from different services may feel awkward or reluctant to interact with personnel from different organisations. Despite there being opportunities to see how these interactions play out in training sessions and the impact this has on cross-service engagement, evaluations of multi-agency trainings tend to rely on personnel surveys that generally assess how well emergency teams feel they collaborate (e.g., JESIP, 2022), or on unstructured observations by trainers who broadly comment on the teams' performances. Surveys often provide subjective perspectives and may not fully capture the nuances of interpersonal dynamics or the subconscious biases that influence team interactions (Philpot et al., 2019). Furthermore, unstructured observations fail to chronologically and comprehensively account for all behaviours across multiple interactants (Philpot et al., 2019), limiting the reliability of this approach for studying interoperability. To date, there is a dearth of work which examines what messages of interoperability look like in these trainings and how interoperability is related to increased inter-agency engagement. More comprehensive, observational methods are needed to fully understand the complexity of these dynamics and to develop strategies that effectively foster true interoperability among emergency service teams.

Aims and objectives

The current research will examine this through a structured, systematic observation of the interactions of individuals from the police, fire and ambulance following multi-agency meetings. Specifically, our initial investigation will focus on analysing the physical proximity of individuals to members of their own service compared to others. Our goal is to determine if there is a significant tendency for individuals to cluster with colleagues from their own service, beyond what might be expected by chance. Furthermore, this study aims to assess the feasibility of measuring behavioural markers of the psychological components of interoperability, as introduced by Power, Alcock et al., (2024). These include (1) trust, (2) identification, (3) goals, (4) communication and (5) flexibility. If able to measure these markers of psychological interoperability within multi-agency meetings, the next question posed is: to what extent does psychological interoperability influence interactions between the different emergency services? This will involve assessing the degree to which trust, identification, goals, communication and flexibility are present within multi-agency meetings and how this influences an individual's bias in terms of proximity, attention, communication and openness toward members of their own group compared to others.

By systematically observing and analysing interactions during multi-agency training exercise meetings, we aim to determine whether the components of interoperability can be quantitatively measured in observational data and whether their presence effectively bridges gaps between different groups, or if natural tendencies towards self-segregation and silo-working prevail. This investigation will provide empirical evidence of the effectiveness of our behavioural codebook and offer valuable insights into enhancing interoperability in emergency services, thereby contributing to more efficient emergency response coordination. In summary, while training and preparedness are vital for improving multi-agency interoperability, existing methods such as personnel surveys and unstructured

observations have limitations in capturing the behavioural dynamics of inter-agency collaboration. Our research aims to address this gap through structured, systematic observation, focusing on the physical proximity and interaction patterns among emergency service personnel. By examining the behavioural manifestations of interoperability components like identity, goals and communication, this study aims to provide deeper insights into fostering effective collaboration across diverse emergency service teams. The findings will not only validate our innovative behavioural codebook but also contribute significantly to the development of more cohesive and efficient emergency response strategies.

Method

Ethical approval

The research received full ethical approval from the Lancaster University – ref: FST-2023-3323-RECR-2.

Data set and initial coding procedure

Data were collected from a large-scale emergency training exercise simulating a building collapse of an underground station, which involved over 4,000 responders, 2,500 casualties, and lasted over four days. The exercise was extensively documented using body-worn cameras, hand-held cameras and static cameras, resulting in a diverse array of footage capturing the intricacies of the operation from multiple perspectives. The initial video corpus contained a total of 339 video clips, totalling 119.62 hours of footage. From this initial corpus, our focus was to isolate the most pertinent videos to our research objectives. The primary criterion for video selection was the presence of multi-agency interactions, specifically those involving *all* blue light services. We also prioritised footage with static camera angles to ensure a comprehensive view of the participants and their interactions, as

this was crucial for accurately capturing the dynamics of the exercise. Further, we sought videos that included meetings where leaders discussed aspects related to interoperability dimensions, as these conversations were central to our research questions. To ensure the quality and relevance of the data (i.e., 'optimal capture' - Nassauer & Legewie, 2021), we conducted a detailed video mapping exercise. This mapping involved a systematic coding process for each clip, assessing variables like runtime, participant visibility, audio-visual clarity and content relevance (see Appendix A for details). Videos lacking in audio-visual clarity or not sufficiently showcasing multi-agency interactions were excluded to ensure the data's quality and relevance to our research objectives.

This stringent selection process led us to a final sample of five high-quality videos. These videos, captured from elevated static cameras, offered a comprehensive view of the meeting room, enabling detailed observation of the interactions among different emergency service members. The footage predominantly featured multi-agency meetings led by senior figures, addressing a larger audience composed of subordinates from various sub-services. Around this meeting, there was a significant time of 'milling', in which individuals had the opportunity to interact freely with one-another in an unstructured way. This 'milling' period provided a unique window of opportunity for systematically observing natural congregation patterns – i.e., who chose to associate with whom and whether there was a tendency for self-segregation beyond what might be expected by chance.

To examine this, we separated each video into three broad time intervals, pre-meeting milling time (n= 2), meeting time (n= 5) and post-meeting milling time (n= 5). We then decomposed each milling period footage (n= 7) into a series of 10-second stills. For example, if a milling period lasted from 3 seconds to 103 seconds, we produced 10 unique stills capturing the room at 3, 13, 23, 33, 43, 53, 63, 73, 83, 93, and 103 seconds,

respectively. We chose 10-second intervals as this provided a balance between capturing sufficient detail and managing data volume. In practical terms, each still served as an anchor point, allowing the research team to observe the surround five seconds of video in either direction and to make their recording. For each still period, the research team assigned all visible participants a ***unique ID*** (e.g., Participant 1) and a letter denoting the ***emergency service*** to which they belong: 'A' = ambulance, 'F' = 'fire', 'P' = 'police', 'O' = other. For each still, we also recorded whether each individual's ***nearest neighbour in physical space*** was of their own organisation '1' or not '0' (Philpot & Levine, 2022). Proximity was judged based on visual cues such as the relative position of feet and the presence of objects, ensuring consistent and objective distance estimation (Hoeben et al., 2021). This recording of congregation patterns and proximity aligns with our objective to understand natural tendencies towards self-segregation, a key aspect of interoperability in emergency services. This allowed us to carry out our first analysis, which ascertained natural congregation patterns during informal gatherings – i.e., who chose to associate with whom and whether there was a tendency for self-segregation beyond what might be expected by chance.

Components of Interoperability and its coding

Operationalising the five components of interoperability

Our next analytical focus was to record the interoperability exhibited in inter-agency meetings (n = 5). Here, we aimed to examine both (i) the feasibility of recording markers of interoperability during inter-agency meetings and (ii) the degree to which these utterances could then impact inter-agency levels of engagement, as displayed in the post-meeting stills. To do this, we first needed to operationalise the five components of interoperability defined by Power, Alcock et al. (2024)—trust, identity, goals, communication and flexibility. Given that these five components primarily represent psychological constructs, their translation

into observable behaviours, particularly verbal ones, presented a novel challenge. Given no prior operationalisations existed, we developed these new measures from the beginning.

We took an ethological approach (Dawkins, 2007) to identify verbal cues of interoperability. To develop these new measures and to keep this measure development stage distinct from the final analysis, we initially selected a subset of three high-quality videos that, while not eligible to be included in our final sample, provided rich verbal information and a diverse range of interactions. This approach allowed us to create a codebook independently of the data used in our final analysis. To construct the measures, the research team adopted both a deductive and inductive iterative approach. Specifically, the five components of interoperability (trust, identity, goals, communication and flexibility) as previously outlined by Power, Alcock et al. (2024) pre-defined how interoperability may be observed. Building upon these definitions, the research team then repeatedly observed the subset of videos and identified, discussed and recorded instances (Pitney, 2004) where verbal behaviours appeared to denote a principle of interoperability. This process aligned with our aim to translate psychological constructs of interoperability into observable verbal behaviours, thereby providing a novel method to assess inter-agency engagement and collaboration. The definitions and criteria for these five variables, developed from this ethological work, are detailed below.

Recording the presence of the five components of interoperability

When coding the final data, since the five components of interoperability relied on verbal information gathered exclusively during the meeting phases, we could not use the method of separating the observations into multiple image stills. Rather, each of the five meetings was observed in its entirety, and each meeting was assigned five global scores, one for each of the five components of interoperability. Scores ranged from 1-6, where 1 = *no*

displays of the interoperability component; 2 = *minimal displays*; 3 = *few displays*; 4 = *some displays*; 5 = *several displays*; and 6 = *many displays*. Scores could also be reduced if there was evidence of adverse verbal behaviours which undermined successful interoperability – for example, in the case of communication, where multiple services may be engaged in conversation, but repeatedly talking over one other.

Measures of the five components of interoperability

We now present the definitions of each of the five components of interoperability (see Table 1), adapted from Power, Alcock et al., (2024) and developed during the measure development stage described above. Full descriptions and development rationale for these measures are available as further supplemental material (see osf.io/28vg7/).

Table 1
Behavioural measures of interoperability

Measure	Definition	Observable indicators
Trust	Trust is the extent to which individuals can rely on and are willing to be vulnerable to and act upon the words, actions and decisions of another individual or group, including cognitive and interpersonal trust.	• Trusting others to take on a task; • Accepting judgement of others; • Accepting assistance from other services; • Physical affiliative gestures, such as light touching or handshakes (see Figure 1).
Shared Identity	Identification with a shared, superordinate group that includes all the different emergency services units.	• The use of inclusive language (e.g., 'we', 'us', 'blue lights'); • Discussions emphasising working together as 'one' cohesive unit; • Physical displays of shared humour and laughter.
Goals	Clear understanding of both the overarching team's objectives and those of individual sub-teams. This involves showing respect for and awareness of one's own goals as well as those of others, and an understanding of how these goals interact.	• Discussions of how one service's objectives might impact and align with those of the other services; • Consideration of how service-specific objectives interplay with broader multi-agency goals; • Consideration of timescales, resources and capabilities across services.
Effective communication	Sharing of relevant and meaningful information among team members in a clear and efficient manner, striking a	• Concise messaging; • Avoidance of specialist language; • Use of closed-loop communications;

	balance between providing sufficient detail and avoiding unnecessary information overload.	• Effective use of visual aids; • Evidence of balanced communication between services that avoid one-sided conversations.
Flexibility	The ability of team members to dynamically adapt to evolving situations. This often involves filling in for each other's roles when someone is overloaded, known as backup behaviour, or deviating from standard procedures for expedited decision-making.	• Filling in for another's role when overloaded; • Offering additional resources or staff assistance; • Adapting to changing timescales and other's needs; • Proactive forward-planning, such as discussing contingency plans or potential future events.

Non-verbal engagement behaviour and its coding

The third step of our coding process was to record non-verbal cues that may indicate level of inter-agency engagement in the post meeting phases. Non-verbal cues may be crucial indicators of inter-agency engagement, affecting elements such as trust and cooperation, essential for effective interoperability (Burgoon et al., 1991; Kurzban, 2001; Carney et al., 2005). The analytical objective, therefore, was to ascertain whether the global measures of interoperability exhibited in the prior multi-agency meeting phase may impact the inter-agency engagement of personnel post meeting. In addition to the previously described 'Nearest Neighbour in Physical Space' measure, we recorded four variables capturing interpersonal engagement, as described below (see Table 2). All engagement variables were coded from observation of each video still and the surrounding five seconds of video in either direction.

Table 2

Behavioural measures of interpersonal engagement

Measure	Definition	Observable indicator
Nearest neighbour	Indicates whether a participant's nearest neighbour in physical space (Philpot & Levine, 2022) is from the same service unit or a different service unit.	Records the nearest individual in physical space. Proximity judged using visual cues, such as the relative position of feet and the presence of objects (Hoeben et al., 2021).
Attention neighbour	Assesses whether the participant's attention is directed toward an	Determined by gaze and body orientation and whether the participant was facing,

	individual from the same service unit or another.	looking at or actively engaging with another person (Clack, Dixon, & Tredoux, 2005).
Speaking neighbour	Records whether the individual that person is conversing with (or last spoke to) is from the same service unit or another.	Based on observation of dialogue and identification of conversation partner (Clack, Dixon, & Tredoux, 2005).
Open gesturing	Assesses whether the participant shows behaviours indicative of engagement, openness and receptivity toward a member of the same or different service unit (see Figure 1).	Determined by whether participants engaged in displays of open body postures, nodding, smiling, behavioural mirroring or animated hand movements (Cuddy et al., 2011; Mehu et al., 2007).
Closed gesturing	Records whether the participant displays behaviours indicative of disengagement or discomfort toward a member of the same or different service unit (see Figure 1).	Included displays of disengaging behaviours, such as crossed arms, fidgeting, appearing uncomfortable or looking away (Cuddy et al., 2011). Note. Closed gesturing occurred to infrequently (n = 7) to statistically analyse.

Note. All engagement measures coded as '1' if directed toward a member of the same service and '0' if toward a different service unit. N/A coded as missing data.

Figure 1

Anonymised Examples of Affiliative Touching (Philpot, Liebst et al., 2022), i.e., Physical Trust (left hand image), Open Gesturing (middle image), Closed Gesturing (right hand image).

Illustrations by Charlotte Betts.

Inter-rater reliability

To establish interrater reliability, 21 of the 103 stills (20.39% of the total sample) were randomly selected for independent coding by three members of the research team. Here, we calculated Gwet's AC_1 coefficients (Gwet, 2014), interpreted using Landis and Koch's (1977) guidelines. Of the 11 measures, six measures achieved almost perfect (0.81-1.0) levels of agreement, four achieved substantial (0.61-0.80) agreement and one (open gesturing) received moderate agreement (0.41-0.60) (for full breakdown table, see supplemental material, Table S1, osf.io/28vg7/). The remaining 82 stills were then divided between the three researchers and single coded. The five components of interoperability were global measures taken across five inter-agency meetings. Given that sample size here was small (n of meetings = 5), we adopted a consensus approach for coding. Here, three researchers collectively observed the training videos, discussed ratings independently, and then reached a common consensus through investigator triangulation (Denzin, 2017).

Analysis plan

The planned analysis is separated into three parts. First, we investigate the physical clustering patterns of individuals from different emergency services. Utilising a total of 102 still images (and the surrounding 5 seconds of video in either direction), we analyse whether members of each service unit are more likely to stand next to their colleagues than what a random distribution would expect. These recordings are made during the "milling" phases in between multi-agency meetings, where there is no expectation for individuals to be grouped within agency. We carry out this analysis using Fisher's Exact Tests and Phi statistics, providing an empirical understanding of natural congregation tendencies. Additionally, a binomial test is employed to assess the likelihood of these patterns occurring by chance, given multiple testing considerations.

Next, we provide descriptive statistics to quantify the frequency of observed interoperability components (Communication, Flexibility, Goals, Identity, and Trust) in the five inter-service meetings. This step is crucial for establishing a baseline understanding of how often these components manifest in real-world training scenarios. While the number of meetings was small (n = 5), which may limit generalisability across settings, the inclusion of multiple individuals within and across meetings enabled sufficient behavioural variance to examine our core research questions.

The final part of our analysis involves a series of multiple logistic regression models. These models are used to examine how the observed components of interoperability in earlier meetings are associated with subsequent intra-group engagement behaviours. Specifically, we assess how the global ratings of these components during meetings influence intra-agency physical proximity, attention focus, conversation dynamics, and open gesturing, as observed in the stills captured after the meetings.

Results

Self-segregation

A total of 102 still images were analysed to determine if individuals from the three emergency services tended to physically cluster with members of their own service unit more than would be expected by random chance. For each image, we observed and recorded the proximity of individuals to others, specifically noting whether they were standing next to members of their own service unit. This observed frequency of individuals standing next to their service unit members was then compared to the expected frequency, calculated based on the overall proportion of each service unit represented in the image.

A series of Fisher's Exact Tests were conducted to assess whether the observed interactions within the same service unit were significantly different from what would be

expected by chance for each of the 102 stills. 26 of the 102 stills showed significant spatial clustering, in which individuals were significantly more likely to be standing next to a person of their own service than expected by chance. For example, Still ID 9 showed a significant preference for same-service unit interactions ($p = .001$), as did Still ID 57 ($p < .001$). Phi Statistics were calculated to estimate the effect size of the observed interactions, with significant Phi Statistics ranging between a small effect (Still 7, Phi = 0.46) to a large effect (Still 11, Phi = 0.84), indicating a notable preference for same-service unit interactions in these specific stills.

To address potential false positives due to multiple testing (Chernick, 2011), a binomial test was conducted. Under the null hypothesis with a standard significance level of 0.05, we would expect approximately 5.1 false positives. The binomial test, comparing the observed 26 significant results against this expectation of 5.1, produced a significant p-value of 0.00017, suggesting that the high number of significant findings is unlikely due to chance.

While a significant number of the Fischer's Exact tests yielded significant results, many did not. These non-significant outcomes may reflect genuine instances of non-self-segregating or be influenced by statistical power. Anecdotally, the scenarios involving non-significant results appeared to involve fewer individuals, suggesting a potential influence on the findings. To further explore this further, a Welch Two Sample t-test was conducted, comparing the average number of participants in stills with significant (M = 13.30, SD = 3.60) and non-significant results (M = 16.0, SD = 2.74). The t-test results noted a statistically significant difference between groups ($t = 3.93$, $p < 0.01$, Cohen's $d = 0.78$), suggesting that smaller sample sizes in certain stills may contribute to non-significant results.

Frequency of observed interoperability and its impact on intra-group engagement

Next, we examined the prevalence of the five interoperability components during meetings and assessed how these occurrences impacted subsequent intra-group engagement behaviour.

Effective Communication

Effective communication (i.e., sharing relevant information among personnel in a clear and efficient manner) was the most frequently observed component across meetings component, with a mean score of 4.60 (SD = 1.14) (see Table 3), indicating that it was typically observed between 'some', and 'several' times in meetings.

Table 3

Prevalence of the five verbal markers of interoperability in joint meetings

Dimension	*M*	*SD*	Min	Max	Range
Effective Communication	4.60	1.14	3	6	3
Flexibility	3.80	0.84	3	5	2
Goals	4.20	1.48	2	6	4
Identity	3.00	0.71	2	4	2
Trust	2.60	1.52	1	5	4

Note. The scale used denotes: 1 = *none*; 2 = *minimal*; 3 = *few*; 4 = *some*; 5 = *several*; and 6 = *many*.

We conducted multiple logistic regression models with robust standard errors (to account for data nesting, Huang, 2016) to examine if the interoperability behaviours exhibited in team meetings influenced the likelihood that individuals stood next to, paid attention to, conversed with, or displayed open gestures toward members of their own service unit during the subsequent milling period. This was to identify whether group-level interoperability scores for multi-agency meetings (e.g., multiple positive indicators of communication, flexibility, goals, identity or trust) influenced how team members interacted

with one another during post-meeting “milling” phases. Specifically, are high interoperability scores associated with decentralised and less siloed interaction post-meeting?

The degree to which effective communication was displayed was not associated with physical clustering toward one’s own service (OR = 1.142, p = 0.286) (for full regression outputs, see Tables S2-S5 of supplemental materials, osf.io/28vg7/). However, it showed a positive association with paying attention toward one’s own service member (OR = 1.440, p = 0.023), suggesting a 44% increase in the odds of focusing attention on a member of one’s own service unit for each unit increase in Communication Score. Communication was also positively associated with the likelihood of conversing with and displaying open gesturing toward one’s own service unit (OR = 1.532, p = 0.04 and OR = 2.427, p = <.001). Here, for each unit increase in Communication Score the odds of conversing or displaying open gestures with members of the same service unit increased 53.2% and 142.7%, respectively.

Goals

Discussions about prioritising, aligning and achieving goals were frequently observed during inter-agency meetings (M = 4.20, SD = 1.48). While discussion of goals did not significantly impact whether a service member’s nearest neighbour tended to be from their own service (OR 0.953, p = 0.728), it did marginally predict attention toward their own group (Odds Ratio = 0.686, p = 0.049), with a 31.4% decrease in the odds of ingroup attention focus for each unit increase in Goals. Goals also significantly predicted a decrease in speaking neighbour bias (OR = 0.531, p = 0.007, 46.9% decrease) and open gesturing toward the same service unit (OR = 0.463, p = 0.006, a 53.7% decrease), suggesting that higher goal alignment can reduce the propensity to exclusively with one’s own service unit.

Identity

Mention of shared identity had a mean score of 3.00 (SD = 0.71), indicating moderate verbal displays of among service teams. Despite this lower frequency, shared identity significantly impacted physical clustering with members of the same service unit (OR = 0.710, p = 0.034), with a 29% decrease in the odds of standing next to a member of one's own service unit for each increase in Identity Score. Identity was also negatively associated with focusing attention on a member of the same service (OR = 0.602, p = 0.027) – an approximate 40% reduction in odds. However, Identity did not significantly impact conversation or open gesturing (OR = 0.952, p = 0.857; OR = 0.737, p = 0.371).

Trust and Flexibility

Trust was the least observed, with a mean score of 2.60 (SD = 1.52 - ranging between 'minimal' and 'few'). Flexibility had a mean score of 3.80 (SD = 0.84) indicating moderate displays of adaptability across services. However, due to high correlations with Communication and Goals scores, Trust and Flexibility were omitted from regression analyses to mitigate multicollinearity issues (Tabachnick & Fidell, 2007) (for further detail, see footnote[1])

Discussion

In major emergencies and disasters, effective multi-agency teamwork across different emergency services teams is imperative (Cabinet Office, 2022, 2023b). Interoperability is key in ensuring that team members from different services have a clear understanding of the multi-team system's structure and needs (Power, Alcock et al., 2024).

[1] Scrutiny of the correlations among the five dimensions of interoperability (Trust, Identity, Goals, Communication, and Flexibility Scores), prior to regression modelling, revealed very strong correlations. Specifically, Trust Score showed very high correlations with Communication Score (Pearson's r = 0.968, p < .001) and Flexibility Score (Pearson's r = 0.966, p < .001), while Goals Score was also highly correlated with Communication Score (Pearson's r = 0.832, p < .001) and Flexibility Score (Pearson's r = 0.720, p < .001).

Successful interoperability further fosters psychological and social connections that are crucial for collective sense-making, problem-solving and decision-making. Despite its recognised importance, the challenge lies in effectively conceptualising and measuring interoperability, particularly given the dynamic and often unpredictable nature of emergency scenarios. Through a blend of qualitative and quantitative observational techniques, we analysed inter-agency operational meetings within large-scale multi-agency training exercises. Our aim was to evaluate whether markers of interoperability could be measured in these contexts, and further whether the components observed in these interactions led to improved cross-service interactions post-meetings.

A central challenge this study addresses is the natural inclination toward in-group bias, where, despite shared objectives, different teams may default to self-segregation (Clack, Dixon, & Tredoux, 2005; Dixon & Durrheim, 2003; Schrieff et al., 2005). This is potentially problematic, as siloed working can impede cross-communication, diminish shared situational awareness and delay a coordinated response (Saunders, 2022; Power & Alison, 2017; Power, 2018). By observing the physical and psychological interaction patterns among police, fire and ambulance services, this study aimed to discern whether there's a marked propensity for different emergency service personnel to physically cluster within agency beyond mere chance. Moreover, the current investigation contributed to existing knowledge by assessing the feasibility of measuring behavioural markers reflecting the psychological components of interoperability (trust, identification, goals, communication, and flexibility) as outlined by Power, Alcock et al., (2024). Through direct observation of multi-agency meetings, we created and tested a behavioural codebook of interoperability behaviour that can be used by researchers and practitioners looking to evaluate

interoperability behaviour via observations. This offers new insights into analysing behaviour to fostering effective collaboration across diverse emergency teams.

Through the analysis of video images from inter-agency operational meetings, we observed notable patterns in the physical clustering of service members. Specifically, a quarter of the still images revealed a significant tendency for individuals to stand next to colleagues from their own service unit, more so than random chance would suggest. While not all images exhibited this pattern—highlighting the variability of interaction dynamics—this observed pattern of spatial clustering is consistent with findings from the literature which highlight the natural inclination towards self-segregation even in neutral settings (Clack, Dixon, & Tredoux, 2005; Dixon & Durrheim, 2003; Schrieff et al., 2005). Such behaviour may reflect a desire for familiarity, safety and predictability – particularly in uncertain or minimally structured environments.

However, our study extends these insights into emergency service settings, indicating that despite organisational efforts to foster integration, such as those described by JESIP (2013, 2016, 2021), ingrained social behaviours continue to challenge these initiatives. This finding is significant in light of the wealth of research suggesting that silo-working inhibits multi-agency effectiveness (Alison et al., 2015; Power, 2018) and the recurrent finding from public inquiries and debriefs that silo-working is a pervasive problem during emergencies (Saunders, 2022; Pollock, 2013). These findings underline the importance of addressing cultural and psychological barriers, as suggested by Power, Alcock and colleagues (2024), who emphasised the need for psychological training programs to enhance collaboration.

We also examined the extent to which verbal markers of interoperability components—communication, goals, flexibility, identity, and trust—appeared during inter-agency meetings. Examples of good communication emerged as the most frequently

observed component with multiple positive examples of effective communication (e.g., avoiding use of acronyms), suggesting an awareness of the importance of open and clear communications across responders. This finding aligns with the intuitive understanding that clear and open communication is necessary in inter-agency settings for the alignment of objectives and actions among diverse teams (Kapucu, 2006; Charman, 2014; Waring et al., 2022). The prominence of clear communication within meetings further aligns with Joint Doctrine of JESIP (2021), which stipulate that open and jargon-free communication is crucial for achieving a shared understanding and collective sensemaking among different emergency services. Establishing a clear awareness of group-level goals was the second most frequently observed interoperability component, where meetings scored highly when team members verbally clarified inter- and intra-agency goals, and developed a shared focus on aligning objectives. Similarly, the focus on shared objectives echoes the importance of mutual understanding and respect in enhancing decision-making (House, Power & Alison, 2014; Kirkham, 2009). These effective communication and clear goals are highlighted as essential for transcending the operational differences that often slow coordinated responses during crises (Brown et al., 2021).

Flexibility, defined as the ability of team members to dynamically adapt to evolving situations, was observed to a moderate extent in meetings, reflecting the need for adaptability in emergency situations. This is important, as research has shown that decentralised approaches to teamwork are important for multi-team emergency settings, wherein participants must have a clear understanding of roles and responsibilities across the multi-team system to provide backup behaviour and support to one another (Brown et al., 2021; House et al., 2014). Surprisingly, identity and trust were the least frequently observed components during multi-agency meetings, despite being identified as important team-level

processes across a wealth of studies (Power, Alcock, et al., 2024; Power, Philpot, et al., 2025; Davidson et al., 2023a, 2023b). The limited observation of identity and trust in our findings may reflect the persistent challenges of building a superordinate team identity across different service units (Davidson et al., 2023a, 2023b). The relative absence of identity and trust lends further support to those who have recommended the critical need for targeted strategies, across training and during operations, to build trust and identification across services (Power, Betts et al., 2024). The limited observation of the psychological interoperability aspects of trust and identity may also speak to difficulties in effectively measuring these more 'internal elements' within the dynamic settings of emergency response. For instance, there is debate in the literature about whether trust is a behaviour or an expectation (Hardin, 2002) which is further complicated by the wide range of methods that have been developed to try and measure this illusive construct (Bauer & Freitag, 2017). This also underscores the importance of data triangulation, to combine behaviour and self-report accounts for a more comprehensive assessment of these critical psychological factors (see Philpot et al., 2019).

Additionally, the study explored how exhibited interoperability behaviours in team meetings were associated with physical clustering, attention focus, conversation patterns and open gesturing in subsequent interactions. Expressions of shared identity, although relatively infrequent in occurrence, were negatively associated with intra-agency clustering and attention. This suggests that when present, cultivating a shared team identity may encourage more cross-service physical co-presence. This finding aligns with work by Novelli and colleagues (2010), which showed that shared social identity can reduce interpersonal distancing, even in experimentally manipulated minimal group paradigms where group distinction is arbitrary. Here it is proposed that a sense of shared identity, or 'we-ness',

fosters psychological closeness, which translates into a greater willingness to physically approach and position oneself near those perceived as ingroup. Increased identity did not, however, translate into significantly more cross-service verbal interactions or open body gesturing. One possibility is that shared identity supports an initial willingness to be physically co-present, but deeper interaction may depend on other factors, such as established communication norms, clearly defined role boundaries or confidence in navigating inter-agency procedures and expectations. These may act as mediators that constrain the translation of psychological closeness into more active engagement. This interpretation is consistent with previous findings, which note that while superordinate objectives can help overcome subgroup biases, they do not automatically lead to improved inter-agency communication (Mitchell et al., 2011). Rather, the impact of shared identity may be limited to initial affiliation unless supported by shared norms, clearly defined roles and structured opportunities for interaction (Michell et al., 2011; Thomas et al., 2010). Thus, while fostering a shared identity can successfully break down physical barriers between different service units, encouraging more complex layers of integration, such as verbal communication and non-verbal openness, will likely require additional strategies, support structures and greater buy in. This underscores the need for multi-faceted approaches in training and operational protocols that not only build and reinforce shared identity but also actively facilitate deeper levels of interaction and understanding across services (Davidson et al., 2023b, Power, Betts et al., 2024). The need for these strategies echoes the recommendations of Power, Betts and colleagues (2024), who advocate that frequent and repeated joint trainings are required to increase trust and cross-service cohesion—though also acknowledging that such training efforts are likely hampered by budget reductions, time constraints and other logistical hurdles.

Greater discussion of team goals was associated with a decreased likelihood of individuals paying attention to, speaking to, and being openly receptive towards members of their own service unit over others. This suggests that a strong focus on shared objectives in inter-agency meetings may encourage broader engagement across different service units, potentially fostering a more inclusive and integrated approach to collaboration among emergency services. This finding aligns with the benefits of multi-agency training highlighted by House, Power, and Alison (2014), where understanding each other's operational procedures and goals, and fostering a culture of mutual respect are emphasised as ways to enhance collaborative effectiveness. Our results suggest that such training programs need to further integrate, wherever possible, discussions that emphasise shared goals to continue to reduce in-group biases (Power, Betts et al., 2024), and that repeated, integrated training exercises which focus on shared goals can significantly enhance the speed and effectiveness of decision-making in major emergencies (Kirkham, 2009).

Communication emerged as a mixed factor impacting team dynamics. While examples of communication were most frequently observed across meetings, it was unexpectedly positively associated with the likelihood of individuals focusing their attention on, engaging in conversation with, and being openly receptive towards members of their own service unit at the expense of others. This indicates that effective communication skills not only facilitate intra-service unit dialogue but may also inadvertently reinforce silos by enhancing the cohesion within specific service units at the potential expense of cross-service collaboration. This finding that communication may inadvertently reinforce in-group preferences is counter to previous evidence of multi-agency working across the 9/11 world-trade center disaster, which found that better communication was associated with more efficient crisis management and resource allocation (Kapucu, 2006). This unexpected

reinforcement of in-group preferences, despite shared communication, underscores the complexity of behavioural dynamics in multi-agency settings and resonates with the position of Thomas et al. (2010) that successful interoperability requires more than a common language and effective information sharing.

In sum, our study introduces and tests a behavioural codebook for assessing interoperability in 'real-time' interactions and sheds light on critical dynamics that influence the effectiveness of multi-agency collaborations. The tendency towards self-segregation, the pivotal roles of communication and goals, and the challenges in fully realising components of trust and identity within multi-agency settings emerge as key areas for future focus. These results highlight the need for more comprehensive, observational methods to fully capture and address these dynamics, as also suggested by Philpot et al. (2019) in their critique of an over-reliance on survey methods for evaluating social interaction. These insights also contribute significantly to the ongoing discourse on enhancing emergency service interoperability (e.g., Davidson et al., 2023b; Power, Alcock et al., 2024), offering empirical evidence to inform strategies aimed at promoting more integrated and efficient emergency response efforts.

Practical implications

In response to Grote and Kozlowski's (2023) call for more explicit policy recommendations to improve teamwork, our findings offer four practical implications for improving interoperability during multi-agency emergency service trainings and operations.

1. Emphasise shared goals in joint meetings

The results of the current study showed that discussions of shared goals were associated with reduced ingroup biases in attention, speaking and open gesturing. This

supports the view that well-articulated shared goals can help diminish working siloes (see also, Alison et al., 2015) and that explicit goal sharing can reduce the risk of agencies translating shared goals into conflicting agency-specific objectives (Power & Alison, 2017). As a concrete recommendation, training leads should begin and end inter-agency meetings with a structured 'goal round' in which each service briefly outlines their primary operational objectives and how these connect to overarching and shared incident-level aims.

2. Structure communication to avoid reinforcing silos

Although effective communication was the most frequently observed interoperability component, it was unexpectedly associated with increased intra-service biases. While good communication is essential for effective joined working (e.g., Waring et al., 2020), it must be structured in a way to that prevents dominance by any single group. Training leads should implement structured turn-taking protocols, rotate service spokesperson roles across services, and introduce cross-agency discussion prompts to promote more balanced and inclusive participation.

3. Foster shared identity through language and action

The use of shared identity inclusive terms ("we", "us" and "blue lights") (Davidson et al., 2022a; Kordoni et al., 2023) was associated with reduced physical self-segregation. However, it did not increase cross-service conversations or openness. This suggests that while inclusive language can encourage co-presence, it alone may not be sufficient to drive more meaningful engagement (see also, Michell et al., 2011; Thomas et al., 2010). Deeper interaction may depend on factors such as established communication norms, clearly defined role boundaries and confidence in navigating inter-agency procedures. We therefore recommend pairing inclusive language with structured collaborative training tasks that promote these foundations – such as joint scenario planning or cross-role simulations that

reinforce shared ‘we-ness’ while also clarifying roles, requiring coordinated action and supporting familiarity with each other’s operational procedures.

4. Embed systematic behavioural observation into multiagency trainings.

While self-report measures provide valuable insights into perceived collaboration levels and satisfaction with trainings, they often miss the subtle, real-time behaviours that shape interpersonal dynamics (Philpot et al., 2019). This challenge in measuring variables relevant to teamwork has been blamed for leading to biased views that ‘team skills’ are ‘soft skills’, as individuals struggle to conceptualise them (Grote & Kozlowski, 2023). Our study demonstrates both the feasibility and added value of structured behavioural observation in multi-agency training settings. We recommend that emergency service organisations embed structured behavioural observation into multi-agency evaluations – either through trained observations using standardised coding frameworks or through systematic video analysis, where appropriate permissions exist. Triangulating behavioural and self-report data will provide a richer, more accurate picture of interoperability, helping to identify interactional patterns that may otherwise be overlooked and strengthening the evaluation and development of joint training exercises.

Study limitations and conclusion

There are also limitations that warrant discussion. Despite the notable instances of spatial clustering in this study, it is important to recognise that not all stills showed significant results. This variability points to the complex nature of team dynamics, where not every situation may lead to self-segregation. The instances where significant clustering was not observed could be reflective of efforts to integrate or, potentially, the influence of smaller group sizes affecting statistical power. A further limitation is that all statistically significant odd ratios in our study demonstrated limited explanatory power (from very small

to small-medium effects in magnitude), which is not uncommon in real-world data analysis where numerous unmeasured variables may influence outcomes (Carey et al., 2023; Götz, Gosling, & Rentfrow, 2022). This limitation underscores the complexity of emergency services' interactions and the myriad factors that can affect interoperability in dynamic situations. Finally, the high correlation observed among some of the five markers of interoperability necessitated the exclusion of certain variables from our analyses to avoid multicollinearity issues (Tabachnick & Fidell, 2007). This limitation means we could not simultaneously assess the full interplay of trust, identification, goals, communication and flexibility within our models. Consequently, the isolated impact of each component may not fully encapsulate the synergistic effect these components have when functioning together in real-world settings, highlighting a gap for future research to explore such interdependencies.

In conclusion, this study provides valuable insights into emergency services' interactions in multi-agency training settings. We developed a behavioural codebook of interoperability for real-time interactions, offering a practical tool for measuring multi-agency collaboration. Our empirical work highlighted the roles of communication and shared goals for fostering inter-agency interactions, while recording the challenges related to identity, flexibility and trust. Further focus on these dynamics will be essential for advancing the interoperability field and for helping multi-agency teams work seamlessly together in times of crises.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to review and improve the language and readability of the final submission. After using this tool, the

authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Alison, L., Power, N., van den Heuvel, C., Humann, M., Palasinksi, M., & Crego, J. (2015). Decision inertia: Deciding between least worst outcomes in emergency responses to disasters. *Journal of Occupational and Organizational Psychology, 88*(2), 295-321. https://doi:10.1111/joop.12108

Allen, D. K., Karanasios, S., & Norman, A. (2014). Information sharing and interoperability: the case of major incident management. *European Journal of Information Systems, 23*, 418-432. https://doi.org/10.1057/ejis.2013.8

Australian Institute for Disaster Resilience (2023). *Australian Emergency Management Arrangements.* *https://knowledge.aidr.org.au/media/10162/handbook_aema_web_2023.pdf*

Bauer, P. C., & Freitag, M. (2017) 'Measuring Trust'. In Uslaner, E. M. (ed.) *The Oxford Handbook of Social and Political Trust* (Oxford: Oxford University Press), pp. 15–36.

Brown, O., Power, N., & Conchie, S. M. (2021). Communication and coordination across event phases: A multi-team system emergency response. *Journal of Occupational and Organizational Psychology*, 94(3), 591–615. https://doi.org/10.1111/joop.12349

Burgoon, J. K. (1991). Relational message interpretations of touch, conversational distance, and posture. *Journal of Nonverbal Behavior, 15*(4), 233-259. https://doi.org/10.1007/BF00986924

Cabinet Office (2022). *The UK government resilience framework.* Available at: https://www.gov.uk//government/publications/the-uk-government-resilience-framework

Cabinet Office (2023a). *Strategic policing requirement.* Available at: https://www.gov.uk/government/publications/strategic-policing-requirement-2023

Cabinet Office (2023b). *CONTEST: The United Kingdom's strategy for countering terrorism, 2023.* Available at: *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1186413/CONTEST_2023_English_updated.pdf*

Carey, E. G., Ridler, I., Ford, T. J., & Stringaris, A. (2023). Editorial Perspective: When is a 'small effect' actually large and impactful?. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *64*(11), 1643–1647. https://doi.org/10.1111/jcpp.13817

Carney, D. R., Hall, J. A., & LeBeau, L. S. (2005). Beliefs about the nonverbal expression of social power. *Journal of Nonverbal Behavior, 29*(2), 105–123. https://doi.org/10.1007/s10919-005-2743-z

Charman, S. (2014). Blue light communities: Cultural interoperability and shared learning between ambulance staff and police officers in emergency response. *Policing and Society, 24*(1), 102–119. https://doi.org/10.1080/10439 463.2013.784306

Chernick, M. R. (2011). *The Essentials of Biostatistics for Physicians, Nurses, and Clinicians*. Wiley-Blackwell. https://doi.org/10.1002/9781118071953

Clack, B., Dixon, J., & Tredoux, C. (2005). Eating together apart: Patterns of segregation in a multi-ethnic cafeteria. *Journal of Community & Applied Social Psychology, 15*(1), 1–16. https://doi.org/10.1002/casp.787

Cuddy, A. J., Glick, P., & Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in Organizational Behavior*, *31*, 73-98. https://doi.org/10.1016/j.riob.2011.10.004

Curnin, S., & Owen, C. (2014). Spanning organizational boundaries in emergency management. *International Journal of Public Administration*, 37(5), 259–270. https://doi.org/10.1080/01900692.2013.830625

Curnin, S., Owen, C. & Trist, C. (2014). Managing the constraints of boundary spanning in emergency management. *Cognition, Technology & Work* 16, 549–563. https://doi.org/10.1007/s10111-014-0285-z

Curnin, S., Brooks, B., Owen, C., & Brooks, O. (2023). Perceptions of strategic decision-making by crisis management teams during exercising: Establishing key dimensions. *Journal of Contingencies and Crisis Management*, *31*(3), 338-352.

Davidson, L., Carter, H., Amlôt, R., Drury, J., Haslam, S. A., Radburn, M., & Stott, C. (2023a). A social identity perspective on interoperability in the emergency services: Emergency responders' experiences of multiagency working during the COVID-19 response in the UK. *Journal of Contingencies and Crisis Management*, 10.1111/1468-5973.12443. https://doi.org/10.1111/1468-5973.12443

Davidson, L., Carter, H., Drury, J., Amlot, R., & Haslam, A. (2023b). Using psychology to improve interoperability: A service evaluation of a novel training tool to support multi-agency working in the UK emergency services. *Qeios Pre-print.*

Dawkins, M. S. (2007). *Observing animal behaviour: design and analysis of quantitative data.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198569350.001.0001

Denzin, N.K. (2017). *Sociological Methods: A Sourcebook (1st ed.).* Routledge. https://doi.org/10.4324/9781315129945

Dixon, J., & Durrheim, K. (2003). Contact and the ecology of racial division: some varieties of informal segregation. *The British Journal of Social Psychology*, *42*(1), 1–23. https://doi.org/10.1348/014466603763276090

Durrheim, K., & Dixon, J. (2005). Studying talk and embodied practices: Toward a psychology of materiality of 'race relations'. *Journal of Community & Applied Social Psychology*, *15*(6), 446-460. https://doi.org/10.1002/casp.839

Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science, 17*(1), 205–215. https://doi.org/10.1177/1745691620984483

Grote, G., & Kozlowski, S. W. J. (2023). Teamwork doesn't just happen: Policy recommendations from over half a century of team research. *Behavioral Science & Policy*, *9*(1), 59-76. https://doi.org/10.1177/23794607231192734

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters.* Advanced Analytics, LLC.

Hardin, R. (2002). *Trust and trustworthiness.* Russell Sage Foundation.

Hillen, M. A., Hoeben, E. M., Kok, R., McHale, C., Sexton, M. M., van der Asdonk, S., van de Bongard, D., Bowman, B., Brown, R., Branje, S., Chorney, J. . . Philpot, R., Portell, M., Yuan, S., van der Veen, C., Noordman, J. (In Press). Guidelines for reporting research

using Systematic Coding of Observed human Behaviour (SCOBe). *Quality and Quanity*. https://doi.org/10.1007/s11135-024-01969-9

Hoeben, E. M., Bernasco, W., Suonperä Liebst, L., van Baak, C., & Rosenkrantz Lindegaard, M. (2021). Social distancing compliance: A video observational analysis. *PloS one*, *16*(3), e0248221. https://doi.org/10.1371/journal.pone.0248221

House, A., Power, N. & Alison, L. (2014). A systematic review of the potential hurdles of interoperability to the emergency services in major incidents: recommendations for solutions and alternatives. *Cognition, Technology & Work,* 16, 319–335. https://doi.org/10.1007/s10111-013-0259-6

Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *Journal of Experimental Education, 84*(1), 175–196. https://doi.org/10.1080/00220973.2014.952397

JESIP (2013) *Joint doctrine: The interoperability framework* (1st ed.). Home Office.

JESIP (2016) *Joint doctrine: The interoperability framework* (2nd ed.). Home Office.

JESIP (2021) *Joint doctrine: The interoperability framework* (3rd ed.). Home Office.

JESIP (2022) *Emergency Services Interoperability Survey.* Skills for Justice.

Kapucu, N. (2006). Interagency communication networks during emergencies: Boundary spanners in multiagency coordination. *The American Review of Public Administration, 36*(2), 207-225. https://doi.org/10.1057/ejis.2013.8

Khorram-Manesh, A., Goniewicz, K., & Burkle Jr, F. M. (2024). Unleashing the global potential of public health: a framework for future pandemic response. *Journal of Infection and Public Health*, *17*(1), 82-95. https://doi.org/10.1016/j.jiph.2023.10.038

Kirkham, B. (2009). CBRN preparedness in a small community: How the Isle of Man has integrated responding and support services. *Journal of Business Continuity & Emergency Planning*, *3*(2), 153–160.

Kordoni, A., Gavidia-Calderon, C., Levine, M., Bennaceur, A., & Nuseibeh, B. (2023). "Are we in this together?": Embedding social identity detection in drones improves emergency coordination. *Frontiers in Psychology,* 14, 1-12. Article 1146056. https://doi.org/10.3389/fpsyg.2023.1146056

Kurzban, R. The Social Psychophysics of Cooperation: Nonverbal Communication in a Public Goods Game. *Journal of Nonverbal Behavior* **25**, 241–259 (2001). https://doi.org/10.1023/A:1012563421824

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Locke, E. A., & Latham, G. P. (2006). New direction in goal-setting theory. *Current Directions in Psychological Science*, *15*, 265–268. https://doi.org/10.1111/j.1467-8721.2006.00449.x

Mehu, M., Grammer, K., & Dunbar, R. I. M. (2007). Smiles when sharing. *Evolution and Human Behavior, 28*(6), 415–422. https://doi.org/10.1016/j.evolhumbehav.2007.05.010

Mitchell, R. J., Parker, V., & Giles, M. (2011). When do interprofessional teams succeed? Investigating the moderating roles of team and professional identity in interprofessional effectiveness. *Human Relations*, 64(10), 1321-1343. https://doi.org/10.1177/0018726711416872

Nassauer, A., & Legewie, N. M. (2021). Video data analysis: A methodological frame for a novel research trend. *Sociological Methods & Research, 50*(1), 135–174. https://doi.org/10.1177/0049124118769093

Novelli, D., Drury, J., & Reicher, S. (2010). Come together: Two studies concerning the impact of group relations on personal space. *British Journal of Social Psychology, 49*(2), 223-236. https://doi.org/10.1348/014466609X449377

Philpot, R., & Levine, M. (2022). Evacuation behavior in a subway train emergency: A video-based analysis. *Environment and Behavior, 54*(2), 383–411. https://doi.org/10.1177/00139165211031193

Philpot, R., Liebst, L. S., Lindegaard, M. R., Verbeek, P., & Levine, M. (2022). Reconciliation in human adults: A video-assisted naturalistic observational study of post conflict conciliatory behaviour in interpersonal aggression. *Behaviour, 159*(13-14), 1225–1261.

Philpot, R., Liebst, L. S., Møller, K. K., Lindegaard, M. R., & Levine, M. (2019). Capturing violence in the night-time economy: A review of established and emerging methodologies. *Aggression and Violent Behavior, 46,* 56–65. https://doi.org/10.1016/j.avb.2019.02.004

Pitney, W. A. (2004). Strategies for establishing trustworthiness in qualitative research. *Athletic Therapy Today*, 9(1), 26-28. https://doi.org/10.1123/att.9.1.26

Pollock, K. (2013) *Review of persistent lessons identified relating to interoperability from emergencies and major incidents since 1986.* Cabinet Office. Available at: https://regulation.org.uk/library/2013-Pollock_Review.pdf

Power, N. (2018). Extreme teams: Toward a greater understanding of multiagency teamwork during major emergencies and disasters. *American Psychologist, 73*(4), 478-490. https://doi:10.1037/amp0000248

Power, N., & Alison, L. (2017). Offence or defence? Approach and avoid goals in the multi-agency emergency response to a simulated terrorism attack. *Journal of Occupational and Organizational Psychology, 90*(1), 51-76. https://doi.org/10.1111/joop.12159

Power, N., Alcock, J., Philpot, R., & Levine, M. (2024). The psychology of interoperability: A systematic review of joint working between the UK emergency services. *Journal of Occupational and Organizational Psychology, 97*(1), 233–252. https://doi.org/10.1111/joop.12469

Power, N., Betts, C., Philpot, R., & Levine, M. (2024). Enhancing interoperability training: Insights from UK emergency responders. [Under Review].

Power, N. Philpot, R, Levine, M., & Alcock, J. (2025). Bridging the principle-implementation gap: Evaluating organisational change to achieve interoperability between the UK emergency services. *Journal of Occupational and Organizational Psychology*, 98(1), https://doi.org/10.1111/joop.70010

Roud, E., & Gausdal, A. H. (2019). Trust and emergency management: Experiences from the Arctic Sea region. *Journal of Trust Research*, *9*(2), 203–225. https://doi.org/10.1080/21515581.2019.1649153

Sanders, C. B. (2014). Need to know vs. need to share: information technology and the intersecting work of police, fire and paramedics. *Information, Communication & Society*, *17*(4), 463–475. https://doi.org/10.1080/1369118X.2014.891632

Saunders, S. J. (2022). *Manchester arena inquiry volume 2: Emergency response (No. 2).* Home Office. https://www.gov.uk/government/publications/manchester-arena-inquiry-volume-2-emergency-response

Schrieff, L., Tredoux, C., Dixon, J., & Finchilescu, G. (2005). Patterns of racial segregation in university

Residence dining-halls. *South African Journal of Psychology, 35*(3):433-443. https://doi.org/10.1177/008124630503500

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon/Pearson Education.

Thomas, K., Bergethon, P. R., & Reimer, M. (2010). Interoperability for first responders and emergency management: Definition, need, and the path forward. *World Medical & Health Policy, 2*(3), 161–166. https://doi.org/10.2202/1948-4682.1070

Waring, S., Moran, J., & Page, R. (2020). Decision-making in multiagency multiteam systems operating in extreme environments. *Journal of Occupational and Organizational Psychology*, *93*(3), 629–653. https://doi.org/10.1111/joop.12309

West, M. A., & Lyubovnikova, J. (2012). Real teams or pseudo teams? The changing landscape needs a better map. *Industrial and Organizational Psychology*, *5*(1), 25-28. https://doi.org/10.1111/j.1754-9434.2011.01397.x

Appendix A

The following variables were measured when sorting raw video data:

- Video duration
- Video date
- Video start time
- Camera type
- Location
- Context (e.g., milling, meeting, operation, walking)
- Agency Present (Fire, Ambulance, and Police, coded as 1 = Present, 0 = Absent)
- Commander Presence (Coded as 1 = Present, 0 = Absent)
- Sound Quality (Coded as 0-3, with 3 being the best)
- Visual Quality (Coded as 0-3, with 3 being the best)
- Inter-group Co-Presence (Coded as 1 = Co-Presence, 0 = Absence)