Psychologists on a Mission: Embracing the controlled chaos of field research

Authors:

David Feltner
Rachel Benton
Heidi Segars
Trevor Patten
Ebernoe Guzman-Bonilla
Susan Simkins
Ericka Rovira
Anne Collins McLaughlin

Precis: We present a case study of designing, running, and iterating a HRI research project in the field. We discuss lessons learned from organizing the moving parts of a multidisciplinary research team in an ambiguous, changing environment of “unknown unknowns.”

Introduction

Operation Clipboard

After jumping out of a plane, crawling through mud, and dodging simulated gunfire, an exhausted cadet finally pushes through enemy territory and finds a moment to rest. Every muscle aches and their throat feels like sandpaper. Up ahead, they spot a canteen and feel relief within reach - until, suddenly another obstacle appears: a fresh-faced researcher blocking the path, clipboard and pen in hand.

“Hi! Could you please fill out this survey?” the researcher asks brightly. With a resigned sigh, the cadet takes the pen. Apparently paperwork is the last obstacle in this mission.

Military Field Training Exercises (FTXs) represent the closest cadets come to replicating combat. Cadets shoot blank ammunition, while other soldiers roleplay an opposing enemy force (OPFOR) and leaders make real-decisions under nearly all the same stressors encountered on battlefields worldwide. While coordinating field data collection with 800 West Point cadets doing

their summer training presents significant logistical challenges, it provides an unparalleled environment for observing how Army leaders incorporate new technology. These contextual characteristics, critical for ecological validity (Kilhstrom, 2021), made this training an ideal environment for investigating how future military leaders will incorporate robots into their tactical operations.

The Army continues to modernize its ground combat forces with robotics to enhance soldier capabilities and survivability (Endsley, 2015; Szegedi et al., 2017). The Department of Defense envisions these technologies improving decision-making on the battlefield (DoD, 2023). Specifically, semi-autonomous robotic reconnaissance “dogs” have the potential to scout the enemy without endangering soldiers’ lives, but further research is required to determine their optimal use.

While the cadets had their directives, we researchers had our own mission: to gather data on human-autonomous teaming in naturalistic settings to improve the design of military technology, decision-making, and tactics utilizing robot teammates. We aimed to explore the attitudes, usage, and initial effects of the robotic teammate across different tactical scenarios. To accomplish this involved navigating fieldwork logistics, coordinating with different echelons of military leadership, and extracting insights from participants on days where they were pushed to their limits.

Importance of Field Research

The complexity of military training environments highlighted the challenges and benefits of conducting field research in high-stakes, dynamic settings. The military remains a complex domain wherein field research can yield the richest data with all contextual factors present (Maner, 2016). Further, field research can yield highly relevant results (Mackie, 1968) with high external validity (Johnson & Baker, 1974). In a FTX, young soldiers balance leadership responsibilities, tactical dilemmas, harsh terrain, all while dealing with a thinking enemy. A laboratory setting just doesn’t capture the nuances of all these contextual factors; however, in a field setting, researchers can observe how soldiers will think and act in combat realistic scenarios.

While field research provides unique insights into real-world decision-making and operational effectiveness, it also can involve significant logistical and methodological challenges. The difficulties of gathering applied data within an operational environment has a long history in human factors research. Fifty years ago, Johnson and Baker (1974) presciently summarized all that we would face in field work:

> “Subjects typically are directly assigned to the test, and test time is also duty or job time. The experimenter is an intruder who has disrupted schedules for reasons which are often not explained or understood. Further, the tasks assigned are sometimes trivial, sometimes difficult, but always a disruption of routine. The result is that the experimenter is confronted with a neutral or even negative attitude and must explicitly motivate subjects.”

Further, Johnson and Baker (1974) identified numerous attributes of field research that differ from lab studies, including that field research is a “messy” problem with many degrees and freedom and requires an “eclectic” methodological approach. Recognizing these longstanding challenges, we meticulously planned our original study. However, in field research, the best-laid plans have to be quickly adjusted when faced with the unpredictability of real-world execution.

The Best Laid Plans

Know Thy User

Executing our study within such a large-scale training exercise first meant understanding the structure of the organization itself and how West Point conducts summer training. We learned military jargon, the task organizations (Figure 1), and about the missions they would train on in the mountains surrounding West Point (Figure 2). Across the fourteen-day FTX, 800 cadets organized into military formations, faced thinking enemies in rugged terrain, and solved complex tactical problems under duress.

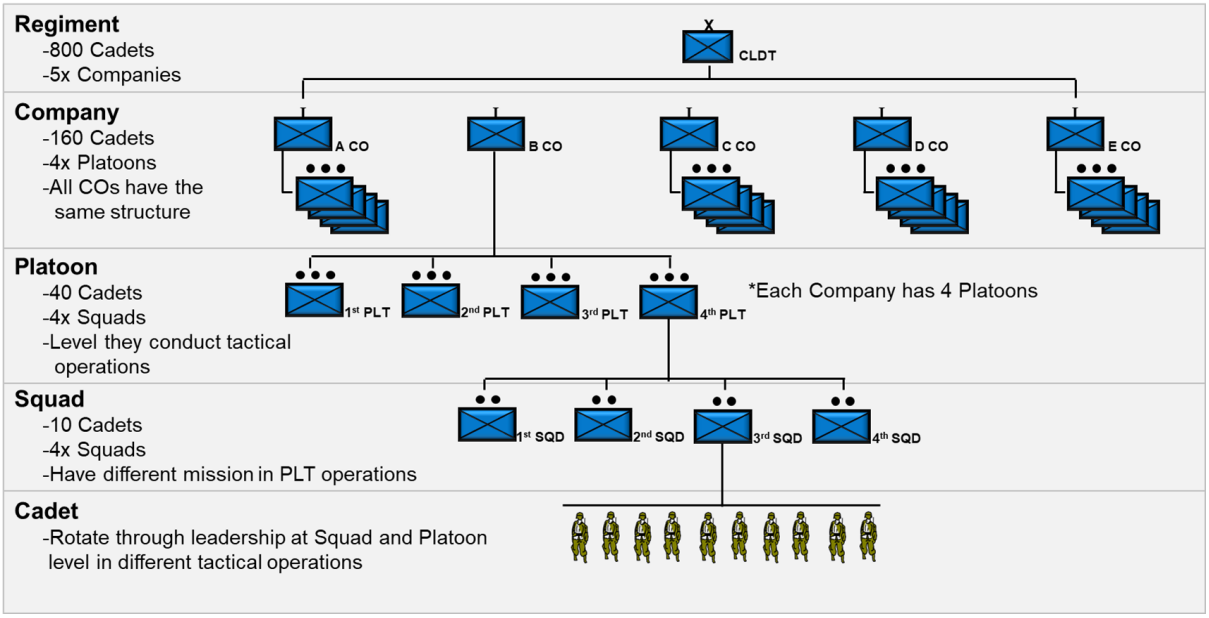

Figure 1. We familiarized ourselves with West Point’s training structure and terminology. Within the 800 cadets out for training, we would have one Platoon come through our lane and use Spot each day. That platoon would be organized into four squads of 10 cadets, each with their own specific task to execute during the lane. This organization informed our planned manipulations and survey distribution plan.

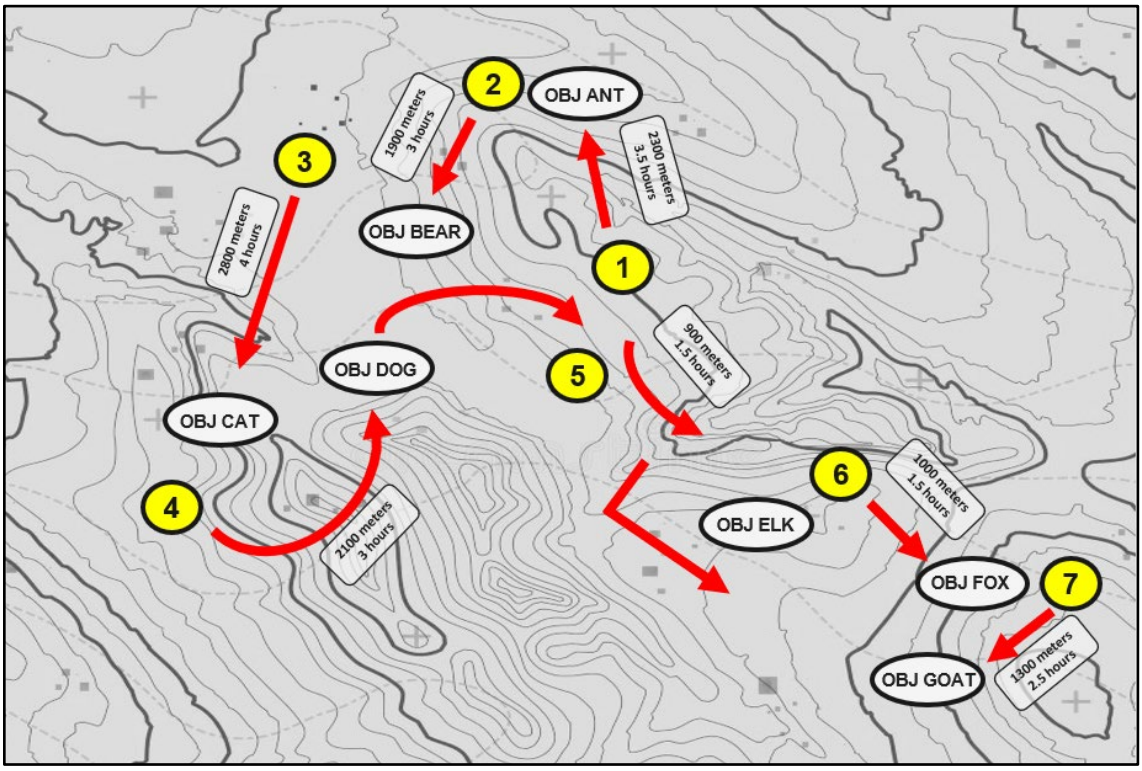

Figure 2. Illustrative topographical map depicting cadet's movement over the 8-day summer training FTX. Except for a helicopter transport from Objective Dog to Objective Elk, all movements were completed on foot through densely forested, mountainous terrain. Researchers hiked with Spot to locations between objectives to meet with cadets.

Extensive planning prior to data collection consisted of three parts: 1) create the research design based on the research questions, cognizant of the structure of the summer training and cadet platoons, 2) data collection logistics, and 3) travel logistics. To prepare, we considered other human factors studies conducted in the field (Fabiano et al., 2022; Fabiano et al., 2019; Woodall et al., 2023). However, in hindsight, past publications—biased toward successful projects—provided little practical guidance on logistics. While prior studies inspired us, they did not prepare us for the inevitable challenges of fieldwork.

We also considered the "lessons learned" literature, particularly a case study published in the HFES Proceedings (Schreck et al., 2023). While this article outlined many of our eventual experiences, its warnings still did not fully prepare us. We followed the advice of this paper, for example, sending one author to the training grounds to understand the environment. (This environment would change greatly due to flooding). We set up an all-team communication channel via SMS chat to synchronize efforts. (Unfortunately, this method left us scrolling through hundreds of daily messages to find information). We replicated poor survey markings (light, outside the bubbles, double marks) in the lab to make sure our software could read those surveys, which it did. (However it could not read markings from the actual surveys due to alignment issues). We chose a hotel close to the data collection site and drove a 4WD vehicle.

(We would not be allowed to use this vehicle due to the flooding). The one piece of advice that we followed most often was their advice to “Improvise, adapt, and evolve.”

Survey Creation

Three subject matter experts from areas within psychology—Teams (Dr. Susan Simkins), Human-Robot Interaction (Dr. Ericka Rovira), and Individual Differences (Dr. Anne McLaughlin), came together to form the core interdisciplinary research group. Each discipline contributed unique research approaches and goals, fostering rich discussions, productive cross-talk, and eventual compromises during preparations for the field research. These diverse perspectives and goals resulted in a comprehensive survey packet consisting of 4 introductory measures, 6 pre-HRI measures, and 10 post-HRI measures. Our design was precise, had clear manipulations, and looked great as we looked at our beautiful plan on the whiteboard (Figure 3).

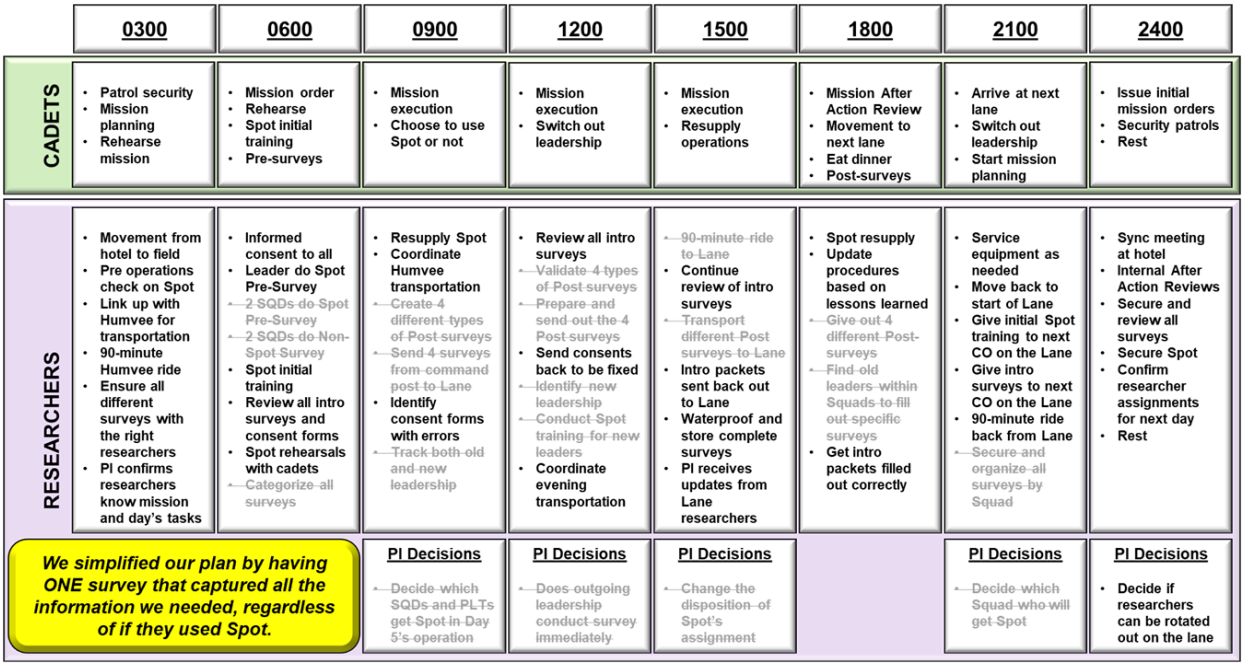

Figure 3. We laid out our tasks and potential decisions to be made throughout the day. In the beginning, we tried to do all of the tasks and realized that we needed to simplify our plan. As we refined our processes, we removed extra steps and became more efficient. Remember the KISS principle - keep it simple.

As the data collection plan took shape, the team gained a deeper understanding of the cadets’ summer training schedule. We held virtual meetings with USMA Plans Officers to understand the nuances of West Point and summer training, ensuring data collection would not disrupt training. They provided insights into the role for Spot during training. With the Plans Officers, we tested Spot’s capabilities on site at the training grounds, making sure that it could navigate stairs, move in buildings, and walk through the woods. After multiple discussions in the months leading up to the training, we identified two training objectives that were best suited to

integrate Spot. Spot would be a part of the “ambush” mission at Objective Ant and the “raid” mission at Objective Elk (Figures 3 & 4).

With measures finalized and plans integrated into the summer training timeline, the team prepared survey packets for the research assistants. Each packet included the appropriate survey versions, administration protocols, and a detailed script to ensure uniformity and accuracy across the expansive training area. To codify our plan we developed a detailed research plan down to the hour for data collection days (Figure 3) - we were ready!

While this process proved valuable, an unanticipated challenge emerged: in reality there were multiple layers of leadership between high-level planners and on-the-ground leaders. The ivory-tower timelines of the planners simply did not reflect reality on the ground, so we adapted.

Planning Lessons Learned

The most prominent takeaway for planning out large-scale field research is to keep the data collection process as simple as possible. Our original research design was methodologically rigorous, created from perspectives inconsistent with reality. For example, some of the individual differences scales involved a pre- and post-interaction measure, which doubled the survey interactions for each task, thus doubling the physical amount of surveys and time commitment. Further, we initially planned to target specific squads within each platoon for Spot integration, incorrectly assuming a level of stability in leadership assignments. Had we designed a simpler, more flexible research plan, we could have avoided many of these complications that arose from physical demands, time constraints, and shifting leadership dynamics.

Figure 4. Left: Spot as controlled by a research assistant in the “raid” mission. Right: Spot as controlled during the “ambush” mission. The researchers controlled Spot, interfacing so Spot could quickly and accurately respond to verbal commands from the cadet patrol leaders.

Once we were in the field, it became clear that real-time decision-making was a mixture of the training cadre, logistical coordinators, and OPFOR leaders who were all working off their own set of priorities. As a result, some of the key assumptions we made based on our initial planning discussions did not hold up in practice. For example, while waiting with Spot to engage

with our next platoon, we were frequently in the wrong place as we were operating off of the original - and now outdated - timetables. Particularly when working in the military domain, we discovered it is extremely important to identify who is responsible for the on-the-ground execution in order to know the “ground truth.”

Time to Execute

Chaos Arrives

Conducting field research in July in the Hudson Valley came with inherent challenges: heat, insects, rugged terrain, and poor cell service. These were compounded by unprecedented weather—a thousand-year rainstorm that brought 3.5 inches of rain per hour, flooding barracks, causing mudslides, and halting operations. This flooding destroyed some roads and made others impassable (Figure 5); these unforeseen challenges demanded we improvise, adapt, and evolve.

Figure 5. Left: The flooding caused significant damage to infrastructure with some roads even completely collapsing on West Point and the surrounding training areas. Right: OPFOR soldiers from Fort Liberty, NC help someone stuck in the rapidly rising water.

The Reality of Logistics & On Site Coordination

Accounting for the flooding’s damage, USMA modified the training schedule and continued the mission with updated safety protocols. Our by-hour plan for survey administration and execution was near useless. Washed out roads meant our personal vehicles were no longer authorized creating several unforeseen problems: we had to coordinate all movements in a military vehicle which restricted our flexibility, USMA needed to issue us helmets to ride in Humvees, commute times increased by an hour each direction, and each researcher had to hand carry all their equipment and surveys rather than storing it in a vehicle. Lastly, most of the time, team members would have no cell service. We held discussions and empowered all

members of the team to exercise initiative, with the guideline to always keep the research question in focus when improvising or adapting and document all adaptation (Figure 6).

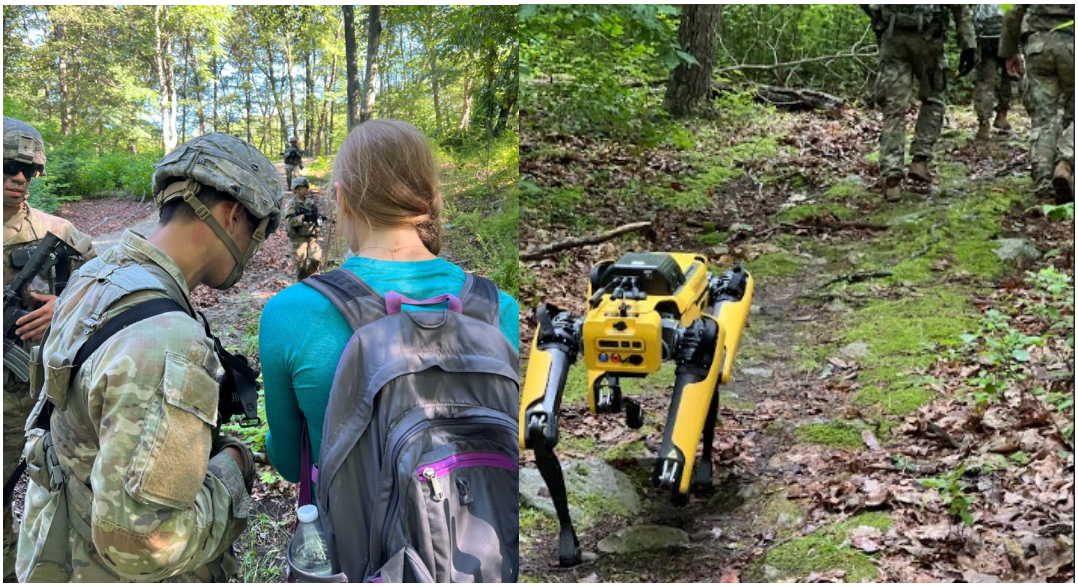

Figure 6. Some cadets independently opted to use Spot as a scout within their formation, utilizing its reconnaissance assets. Others were uncertain about its role and opted not to include it in their operations.Left: The platoon leader opts to use Spot for reconnaissance and watches on the research assistant's tablet. Right: Spot during the "reconnaissance" mission.

With the original plan no longer feasible, the team sought out leaders on the ground who were facilitating the training. Eventually, we found the person with the knowledge, transportation resources, and authority to help us - the OPFOR Company Commander. Charged with maintaining the training throughput, he synchronized our timeline across all the different entities and passed that information to all. We became an additional coordination for him, which had us showing up several hours early just to get arranged rides and updated on the day's schedule, which changed by the minute. His aid ensured that our researchers were at the right place at the right time, although we needed to stay prepared to move at a moment's notice (Figure 7).

Figure 7. The West Point and USMA researchers collaborated with USMA and OPFOR leaders.

Surveys in the Field

Based on our planning, we scheduled introductory survey times with all platoons about to enter the FTX mission with Spot. The plan for this first interaction was straightforward: gain consent to participate in the research, administer the introductory surveys, and introduce the cadets to Spot and its capabilities. Researchers would provide a standardized description of Spot's capabilities, along with suggested ways to effectively deploy Spot.

While the plan seemed simple, its execution proved anything but easy. Leadership roles frequently changed throughout the FTX at the discretion of the individual officers in charge, making it impossible to assess leaders prior. To adjust, we opted to distribute the surveys to all cadets. While this approach increased the risk of survey fatigue, it guaranteed that the relevant leadership positions' experiences with Spot were captured. Another significant challenge was the sheer number of competing requirements that impacted our scheduled survey time. Recognizing that the cadets' primary goal was to rehearse and succeed on their daily mission, we adjusted their instruction to fit around other events while still adhering to the script (Figure 8). However, this reality was different from what we had envisioned during our planning meetings.

Figure 8. Survey completion competing with dinner time in outdoor weather.

Reflecting on our experience, there are a few adjustments that could improve flexibility during execution. We should have included a filtering question in the survey to find the leaders during the mission. This approach would allow us to identify leadership positions during data cleaning and analysis, mitigating the issue of tracking constantly rotating leadership. In the future, we could coordinate with summer training planners to be driving via Humvee for specific times and days. This request could be synchronized through the unit's operations cell and incorporated into the daily *Logistical Sync* meeting. By implementing these changes, we could enhance our adaptability and ensure smoother execution in this environment.

Figure 9. Before and after. Left: Researchers dressed in required helmets for riding in a HumVee to the wooded area where Spot would be used by cadets. Middle: A break for dinner before collecting introductory surveys for the next day from cadets. Right: The long hours of

data collection and patrolling with Spot in the mountainous terrain took a heavy toll on the researchers. Research assistant after he developed the “thousand yard stare” on Day 3.

Discussion: Mission Accomplished

Intellectual Merit

The Army continues to incorporate new technologies on the battlefield, and our work provides a deeper understanding of the relationship between human and robotic teammates in military settings. Hundreds of future military leaders interacted with a robotic teammate and we collected thousands of surveys to gain insight into that relationship. By integrating our research team and the robot into a realistic military FTX, we gained novel insights and paved the way for future research with West Point where we could test new technology as it becomes available.

To ensure our experiences and lessons learned are effectively shared, we are disseminating our findings through multiple publications. We gained significant knowledge about HRI in field conditions, as well as the current limits of robotic teammates for military applications. These findings are part of a magazine article designed to communicate the practical implications of robots in military operations to practitioners in the field (under review, *Modern War Institute*). Additionally, we are submitting a journal article exploring individual differences in how teams perceive and interact with robots. A third paper was motivated by the challenges we faced in data processing and analyses. When circumstances forced us to abandon digital surveys, we reverted to large-scale paper survey distribution and collection. This experience provided valuable lessons about handling and digitizing massive amounts of paper survey data, leading us to submit a paper detailing our process for converting paper responses into electronic form (under review, *Behavioral Research Methods*).

Broader Impact - Building Future Researchers

Despite the heat, the floods, and all the challenges, our team not only survived but emerged stronger, gaining invaluable practical knowledge about field research. Both our undergraduate and graduate student cohorts gained intense research experience and unique field insights. Notably, two undergraduates have since graduated and are now pursuing doctoral degrees in human factors. Additionally, one teammate was so inspired that he is now serving as a Postdoctoral researcher at West Point. Although the research team is now slightly dispersed, this shared “crucible experience” will remain a foundation they can draw upon, providing lessons and skills that will serve them well in the future.

Take Home Messages

- **Keep the psychological question in focus.** All adaptations and changes we made were in service to our research question: what are the team and individual differences that influence perceptions of HRI before and after interacting with a robot teammate?

- **Invest time and effort in building and maintaining team mental models:** Shared understanding of a plan enables flexibility when obstacles, plans shift, or unexpected 1,000 year rains flood the entire region.
- **Preparation and development are key.** The front end effort to build the research team through mentorship, self-development, and research opportunities within the team combined with good team mental models allowed research assistants to take initiative, especially when project leaders could not be contacted.
- **Keep the project simple.** Overly complex research designs and data collection plans should be avoided. For example, collecting surveys from all participants was more simple than isolating the specific squads that could have interacted most with the robot. In the end, this decision was beneficial to our understanding of pre-study individual differences in cadet perceptions of robots as we had data from hundreds of participants.
- **Invest time in After Action Reviews:** Because we travelled over an hour each evening, usually in the same car, back to the hotel, these after-action reviews happened naturally at first. We then formalized them into meetings where we took notes and preserved the experiences and decisions of each day in the field. A paper trail of decisions will be invaluable later.
- **Meet the right people**. As the time approaches, identify and work with a liaison with appropriate knowledge and power who will be physically present during the training and collection.
- **Real-time Organized Communication Procedures:** Choose a tool, structured beyond a simple group chat, that succinctly provides shared understanding across all team members. Ideally, this tool works offline but communicates anytime a cell signal is found.

Conclusion

To prepare for this field study, we consulted previous literature, coordinated with the people on the ground, and tried to think through every eventuality. And while our attempts were admirable, we still were not prepared and we don't think that we could truly have been prepared for every eventuality. When chaos inevitably strikes, we learned that mission success comes down to individual researchers understanding the intent of the mission, the nuances of the research question, and working together as a team to improvise, adapt, and evolve.

Figure 10. Research teams from USMA and NC State partnered to tackle an incredibly difficult field data collection.

As future researchers go out into the field, we recommend you prepare the best way that you can. Develop the research acumen of your team, think through problems, keep your approach simple, and always be ready to improvise, adapt, and evolve. Good luck out there!

Sidebar # 1

Mission Command

The mission command philosophy is the U.S. Army's approach to command and control. It empowers subordinate decision-making and decentralized execution through mission orders, enabling disciplined initiative to achieve the commander's intent. Commanders rely on the innovation and decisive action of subordinates to operate effectively in complex environments. This approach balances the inherent risks of decentralized operations with the seven principles of mission command: competence, mutual trust, shared understanding, commander's intent, mission orders, disciplined initiative, and risk acceptance. Mission command applies not only to soldiers on the battlefield but also to research teams operating in challenging conditions—such as being spread across West Point with no cell phone service. While all seven principles are vital, the following sections highlight a few of these tenets and illustrate how they worked together to make a complex field data collection effort possible.

Shared Understanding and Commander's Intent

Before the mission begins, commanders—or the Principal Investigator (PI) in this case—are responsible for clearly communicating their intent down to the lowest level. Everyone must understand the mission's purpose, the key tasks to complete along the way, and the desired

end state. A common saying is, *"No plan survives first contact,"* so a well-articulated intent allows leaders at all levels to adapt and adjust plans as needed to achieve the goal. In this research effort, the PI had over 20 years of experience, but she couldn't be everywhere at once. Every researcher had to be thoroughly trained to understand the processes and procedures at every stage. When unexpected challenges arose and the intent shifted, the PI had to make difficult decisions and ensure changes were clearly communicated across the entire team. This shared understanding and alignment with the commander's intent enabled the research team to adjust effectively, mirroring how mission command works in Army operations and research environments alike.

Competence and Disciplined Initiative

For both inexperienced researchers and soldiers to make effective on-the-spot decisions, they need a solid foundation of knowledge and training. Achieving competence requires a combination of education, training, and self-development. The Army builds competence through progressive individual and collective training, where teams solve increasingly complex problems to enhance their skills. Similarly, this research lab developed competence through rigorous academic coursework, collaborative efforts between undergraduate and graduate students, and ongoing mentorship. These experiences built trust and confidence within the team, ensuring that decisions made in the field maintained the integrity of the research.

Conclusion

Mission command does not happen by accident; it must be practiced consistently and intentionally. By focusing on the seven principles of mission command, leaders build trust with their teams, enabling them to operate effectively even in challenging conditions. With a foundation of competence, leaders can share a clear vision of their intent and trust their subordinates to take disciplined initiative to accomplish the mission. Whether on the battlefield or in a research lab, applying the mission command philosophy ensures teams are prepared to overcome any future challenges—whether in combat or in field research.

References

Ajzen, I. (1991). The Theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-T

Department of the Army. (2019). *ADP 6-0: Mission Command: Command and Control of Army Forces.* U.S. Army. https://irp.fas.org/doddir/army/adp6_0.pdf

Endsley, M. R. (2015). *Autonomous horizons: System autonomy in the Air Force—A path to the future (Volume I: Human autonomy teaming)*. US Department of the Air Force. http://dx.doi.org/10.13140/RG.2.1.1164.2003

Fabiano, B., Pettinato, M., Currò, F., & Reverberi, A. P. (2022). A field study on human factor and safety performances in a downstream oil industry. *Safety Science, 153*, 105795. https://doi.org/10.1016/j.ssci.2022.105795

Fabiano, B., Pettinato, M., Reverberi, A. P., & Currò, F. (2019). Human Factors and Safety Management: A Field Study on Safety Performance in the Process Industry. *Chemical Engineering Transactions, 77*, 283–288. https://doi.org/10.3303/CET1977048

Kihlstrom, J. F. (2021). Ecological validity and “ecological validity”. *Perspectives on Psychological Science, 16*(2), 466–471. https://doi.org/10.1177/1745691620966791

Maner, J. K. (2016). Into the wild: Field research can increase both replicability and real-world impact. *Journal of Experimental Social Psychology*, *66*, 100-106. https://doi.org/10.1016/j.jesp.2015.09.018

Molendijk, T., & Kalkman, J. P. (2023). Towards a soldier-based view in research on the military: an empathetically critical approach. *Social Sciences, 12*(2), 51. https://doi.org/10.3390/socsci12020051

Rabideau, G. F. (1964). Field Measurement of Human Performance in Man—Machine Systems. *Human Factors, 6*(6), 663-672. https://doi.org/10.1177/001872086400600607

Schreck, J., Sikorski, E., & King, M. (2023). I thought we tested that!: Lessons learned from a team communications field study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 67*(1), 86-90. https://doi.org/10.1177/21695067231194321

Szegedi, P., Koronváry, P., & Békési, B. (2017). The use of robots in military operations. *Scientific Research & Education in the Air Force, 19*(1), 221-230. http://dx.doi.org/10.19062/2247-3173.2017.19.1.25

Deputy Secretary of Defense Kathleen Hicks Announces Publication of Data, Analytics and AI Adoption Strategy. (2023, November 2). U.S. Department of Defense. https://www.defense.gov/News/Releases/Release/Article/3577857/deputy-secretary-of-defense-kathleen-hicks-announces-publication-of-data-analyt/

Woodall, J. L. A., Sak, J. A., Cowdrick, K. R., Bove Muñoz, B. M., McElrath, J. H., Trimpe, G. R., Mei, Y., Myhre, R. L., Rains, J. K., & Hutchinson, C. R. (2023). Repetitive low-level blast exposure and neurocognitive effects in army ranger mortarmen. *Military Medicine, 188*(3-4), e771–e779. https://doi.org/10.1093/milmed/usab394