

CONVERSATIONS ON STRATEGY

PODCAST
TRANSCRIPT

Paul Lushenko, Rob Sparrow, and Adam Henschke AI, Trust, Culture, and the Military

Contesting Paul Scharre’s influential vision of “centaur warfighting” and the idea that autonomous weapon systems will replace human warfighters, this podcast proposes that the manned-unmanned teams of the future are more likely to be minotaurs, teams of humans under the control, supervision, or command of artificial intelligence. It examines the likely composition of the future force and prompts a necessary conversation about the ethical issues raised by minotaur warfighting. The guests also explore culture and trust in relation to AI and the military.

Keywords: manned-unmanned teaming, centaur warfighting, artificial intelligence, future force, ethics, trust

E-mail usarmy.carlisle.awc.mbx.parameters@army.mil to give feedback on this podcast or the genesis article.

Stephanie Crider (Host)

This is part one of a two-part podcast. You’re listening to [Conversations on Strategy](#).

The views and opinions expressed in this podcast are those of the authors and are not necessarily those of the Department of the Army, the US Army War College, or any other agency of the US government.

My guests today are Assistant Professor Paul Lushenko, PhD; Professor Rob Sparrow; and Assistant Professor Adam Henschke. Lushenko is the director of special operations and a faculty instructor in the US Army War College’s Department of Military Strategy, Planning, and Operations. Sparrow is a professor in the philosophy program at Monash University, Australia. Henschke is an assistant professor in the philosophy section at the University of Twente, Netherlands.

About a year ago, Sparrow and Henschke coauthored “[Minotaurs, Not Centaurs: The Future of Manned-Unmanned Teaming](#).” You can find it in the Spring 2023 issue of *Parameters*. Look for volume 53, issue 1. Lushenko and Sparrow took this concept a step further with their piece, “[Artificial Intelligence and US Military Cadets’ Attitudes about Future War](#).”

Let’s start with a quick recap of your *Parameters* article, Rob and Adam.

Robert Sparrow

Most people who’ve thought about the future of the use of unmanned systems in war over the last decade or so have looked towards a future where robots were used to enhance the warfighting capacity of human beings. They have this vision that you’ll still have human warfighters, but they’ll be surrounded by clouds of autonomous or semiautonomous systems that will advance into enemy territory, deliver fires, [and] enhance the power of the warfighter.

Paul Scharre, who’s [an] influential figure in this area, suggested that the best way to understand this was as what he called centaur warfighting. So, the centaur is a mythical creature with the body of a horse and the upper torso of a man. And that image implies that human beings are going to be in control of the manned-unmanned team.

Adam and I suspect that things might go quite differently. We think a better model is actually the minotaur, another mythological figure, with the head of a bull on the body of a man. And that image draws your attention to the possibility that it might well be the (artificial intelligence or) AI in control, that you might actually have a future in which artificial intelligences are delivering orders to human beings and human beings do the so-called grunt work, that moving across terrain and placing weapons, firing weapons, loading weapons might all be done by human beings at the command of an AI.



Paul Lushenko, Rob Sparrow, and Adam Henschke AI, Trust, Culture, and the Military

And the reason why we think that's likely is because we look to what's happening in civilian life with a system like Uber, for instance, which is not robot taxis operating at the command of the human customer, but actually human drivers doing the biddings of an algorithm. And for various reasons, the physical world is really hard for robots. We've seen vast and impressive improvements in the capacity of AI much more than robotics. And so, we think, in the future, it may well be AI is in control, in significant parts of wars of the future.

Host

Paul and Rob, tell us about your work and its findings.

Paul Lushenko

I came at this question by way of semiautonomous drones, which for many years constituted the bulk of my research, and really understanding public perceptions of the legitimate use of force through these capabilities. But yet, we're told that AI, artificial intelligence, is game-changing.

There's two research questions that I think are embedded within this paper. The first is, how do we really understand the implications of AI for the future trajectory of warfare? People have talked about this being game-changing, yet we don't really understand the implications for decision making as well as oversight, whether it's through human oversight or a degree thereof, if not machines.

And then the second question—which is really important for where we sit today, which is the US Army War College in Carlisle—is the degree to which soldiers, service members generally, will trust partnership with these capabilities. This is the notion of human-machine teaming, which is designed to capitalize on machines to optimize performance on the battlefield. And I think the intended benefit here, in terms of what senior leaders envision for human-machine teaming, is that we would shorten the sensor-to-shooter timeline between ourselves and the adversary. By bringing to bear machine learning and crunching of numbers, so-called big data, we can identify a target and levy an effect quicker than our adversaries, and in the context of great-power competition and conflict, god forbid, this is a comparative advantage.

And so, as I started dialoguing and talking with Rob, the real question that struck me as unanswered, empirically at least, was, is it a foregone conclusion that service members will trust partnership with these capabilities? Second of all, what are the implications of generational differences across the military ranks for levels of trust in human-machine teaming?

And this is pretty consequential because you have some scholars and a body of literature that has consigned junior officers, as well as those training at any number of our service academies or at the Reserve Officers' Training Corps—so those training to become commissioned officers—it consigned them as digital natives. And so, we're really attempting to address these assumptions within this paper.

The paper, in short, is about the research question of trust among junior officers in waiting, who we'll commission soon from the Reserve Officers' Training Corps. And we conduct a really unique survey experiment among an even rarer sample of cadets across the United States. And if I recall correctly, I think our respondent pool consists of about 474 cadets across the United States. So, [it is] much more geographically and demographically diverse than you would get in the existing literature that attempts to tap into the attitudes of these officers in training.

And what we find is a really unique set of findings that dialogues with previous research that I have published on what shapes senior military officers' levels of trust—those at the war colleges. So first and foremost, we find that yes, in fact, cadets can be more trusting of AI-enabled military technologies. But at the same time, they hold more conservative or skeptical views on the appropriate use and oversight of these capabilities.

The second finding is, we find much more alignment in attitudes of trust and support among cadets than we do senior military officers who took the same exact survey three or four months ago. And this is really important because, on the one hand, I found previously that officers at the senior level can actually support capability integration

Paul Lushenko, Rob Sparrow, and Adam Henschke AI, Trust, Culture, and the Military

and partnership but not actually trust it.

And trust, in this context, we define as an expectation that a capability will be used, reliably, as marketed, towards shared goals. And so, in a military context, you would imagine this could look like closing with [and] destroying the enemy. On the other hand, we find that cadets don't necessarily have a disagreement in these attitudes, which leads us to believe, if we take a look at the data, there's a greater fluency [and] socialization with these capabilities.

However, the final finding, I think, tracks closely with what we would anticipate for cadets who are junior in age. They're undergoing intellectual growth, they're undergoing acculturation into an organizational culture that prides itself on hierarchy and certain standards of conduct. We find much more variation in attitudes among cadets than we do for officers, in terms of what we identify as instrumental, normative in operational considerations. [It is] a fear that if we don't adopt a capability, another country will, which increases levels of trust. A belief that if this capability could be abused by leaders, which dampens levels of trust, as well as a felt moral obligation to use it abroad, which can actually increase trust. And so, there's a conflict between these different moral logics that govern overall attitudes.

And then finally, and I think this has room for much more research, and I'd really appreciate Rob's thoughts here as well, is that operational experiences exercise the most important shaping effect for attitudes of trust. So, among this junior population, if they have had prior military experience, which some have (to include, especially, combat deployments), that will increase their level of trust greatly. And so, it communicates to us that there is an important socialization effect that's taking place in the context of military service, even at low doses, that may actually be more consequential than theories of neuroplasticity, which says that your brain is malleable based upon experiences that can sharpen your level of judgment.

So in short, the question we confront is trust among cadets for partnership in these capabilities. We find they can be more trusting, but yet hold conservative views about appropriate use. They don't have a trust paradox like senior officers, and there's a lot more variation due to their young age, intellectual growth, and ongoing acculturation than we see with senior service members.

Rob, did you have anything to add there?

Sparrow

I'm fascinated by this research. Paul has conducted the survey, and I've come in to help with the analysis and some of the implications. I think they're a really interesting set of results.

I think it's important to recognize that working out how much people trust AI in various use scenarios doesn't settle the question of whether or not that trust is justified. People can get AI badly wrong, and it's really important to recognize that we have tendency both to over trust and to under trust AI. People will often develop an exaggerated sense of the capacities of automation. And this has been a real problem for driverless vehicles, for instance. They work most of the time—and so, you go to sleep or you get drunk—and you trust that the vehicle is going to continue avoiding obstacles. And then in that rare scenario where it doesn't work, you're in trouble because you're not capable of taking control of the vehicle in time.

So, that would be an example of over trust in a system. But you also see the opposite problem, where people turn these systems off because they give them results that they don't like, or they insist on checking the outputs of AI, and that slows down—that vitiates—the advantages in terms of speed and processing capacities of those systems.

So, things could go wrong both ways. People could trust them too much; they could not trust them enough. It's really interesting to look at what levels of trust, at the moment, keeping in mind that most people actually don't have much experience with these systems. And so, this is certainly work that will need to be done again in the future.

When we see people who've fought alongside these systems, now how do you feel about them? And then there's still this further question about whether the attitudes of service members are actually appropriate or right.

Paul Lushenko, Rob Sparrow, and Adam Henschke AI, Trust, Culture, and the Military

Adam Henschke

Just to step in there, some of my interest on this, I'm part of a project that's looking at trust and connected vehicles, which is similar to autonomous vehicles.

My part of the focus in this project is on those questions that Rob was raising. Is the component, vehicle, tool, and the system worthy of trust? Paul's engaging in some real interesting work on looking at the trust attitudes that people have. But that's only half the story, and the other half is, are these systems, are these components, and are these tools—whether it's a weapon, a vehicle, or something else—are they worthy of that trust?

And this is, in part, an ethics question. There's other aspects to this as well, there's legal aspects, [and] there's probably political aspects. This is a really, really important area when we're talking about trust, particularly with new technologies that we are not yet properly familiar with because we don't have the proper experience with [them]. Should we be placing our trust in these things?

And in some situations, as Rob was saying, there is the danger of over trust, but there's also a considerable problem of under trust. And, particularly when we're thinking of this in military context, we have to be careful about over trust. But also, if these systems are going to either improve, say, survivability for your own military, or they're going to improve the decision making such that the decisions are actually better, either ethically or pragmatically, we want to ensure that these systems are not under trusted either. Otherwise, that kind of defeats the purpose of them. So, there's some really important questions here, on the attitude side of things, but also on the worthiness of those trust attitudes.

Host

What are the implications of AI for the future trajectory of war?

Sparrow

I'm not sure that anyone really knows the answer to that question, I think, in part, because war is a chaotic business at the best of times, and these are systems that people often don't understand. And in the future, there'll be systems of systems. You'll have lots of AI systems all interacting with each other, and quite how that works out, that's a difficult thing to war game, particularly when it's adversary capacities that will play a role here.

I think it's probably fair to say that it will speed war up. I mean, we're already seeing that, that decisions need to be made more quickly, partially because the munitions are moving at higher speeds—there's swarming munitions. Targeting has become much more precise as sensors have improved. And so, people need to respond more quickly. So, our tempo of conflict will increase.

I do think there'll be real pressure to use these systems, out of a fear that one will be too slow, out of, maybe, over trust. If you think about what happens with chess and go—if you're a professional chess or go player and the best computing system tells you what the best move is, you really need to be thinking long and hard about whether you're going to contradict that.

It's really interesting to think, what if that's true for wargaming in the future? What if AI systems actually can outcompete human generals at wargaming? It's going to be very hard to now insist on human judgment, even though at another level, there's good reasons for wanting war to remain a human activity under human control. But I'm sure Paul and Adam have thoughts on this as well.

Lushenko

I think Rob is correct. And I think it's really a stroke of humility that we admit that we're not really sure what could transpire from the adoption of AI-enabled capabilities at scale and across domain and across the Joint force. But yet—as at least a political-social scientist who's about theory building and theory testing through these unique, but not without fault, empirical techniques—what I attempt to do in this research agenda is typologize the potential outcomes of AI in future conflict.

Paul Lushenko, Rob Sparrow, and Adam Henschke AI, Trust, Culture, and the Military

Rob and Adam have already talked about two useful frameworks here, in terms of centaur and minotaur warfare. And my contribution is to take this to a higher level of abstraction—which is to say the strategic level of warfare—which is about integrating ends, ways, and means to achieve overall war outcomes that support political objectives.

And if we do that, it's not just about centaur and minotaur warfare, which is a rheostat of human oversight at the tactical level of warfare—so think battles and engagements between adversaries—but also about decision making at the strategic level of warfare for things like net assessment [or] what we would call decision-support algorithms. Or, how we optimize limited resources to achieve resilient logistics, which is a key point of integrated deterrence within the (United States Indo-Pacific Command or) INDOPACOM theater of operations, at least from a US military perspective.

If that's correct, then we ought to think about AI-enabled warfare in terms of mosaic or hyperwarfare, as well as what other people have called the singleton, or AI-general warfare. And so, for me at least—this is probably not correct but it's a first stab—it seems that there are four potential outcomes for AI in the future on the battlefield: centaur, minotaur-enabled warfare; mosaic hyperwarfare, which is about strategic decision making with broadly human oversight; and then this almost sci-fi thought on an AI general, or singleton, that would conduct decisions in such a way that is an extreme form of minotaur warfare.

And so, I think this is useful because it takes a lot of the conjecture that's taking place on the beltway right now from the Center of New American Security (Center for a New American Security), the Brookings Institute (Brookings Institution), or whatever, and allows us to derive testable hypotheses in terms of attitudes of trust, which I think is a really good way to understand the broader context for where the research that Rob and I are conducting, as well as Adam, derives from.

Sparrow

I think we should also be open to the possibility that it will fail and fail badly. If [in] a big organization, someone offers to upgrade your [information technology or] IT system, you should be really cautious about that because sometimes those projects, they're almost all inevitably, massively over budget. And sometimes the system that results is actually less capable than the legacy system.

I think it's unlikely that it'll always fail and there's probably some circumstances where it's already more reliable. But replacing the manned submarine fleet with fully autonomous, unmanned submersibles, there's some good reasons to do that. Submarine warfare is precisely the kind of environment where arguments about getting rid of dirty, dull, dangerous roles have a lot of force, but that would be a big gamble and you'd need to be really careful that all of your boats didn't miss the conflict or suffer some sort of terrible IT crash.

There's also historical examples. There's a very famous example called “the flash crash” in the stock market with AI share-trading algorithms. This is a case where you have already got something like strategic decision making by AI systems. And the interaction between those systems actually dropped prices dramatically [and] caused a stock market crash for reasons that were very hard to understand.

So, there's a lot of uncertainty here. I'm on board with the typology that Paul set out, but I think we should also keep in mind the possibility that, along the way, there might be some pretty dramatic failures.

Henschke

To pick up Rob's point, I think we will have to plan for and anticipate failure and probably system-level failures.

This seems maybe not inevitable, but highly likely, that some sets of system-level failures are going to occur as part of the introduction of AI.

And for me, one of the interesting things is, and one of the challenges is, how to go through that so that your system is better in a competitive environment. If you're testing things and testing these sorts of AI systems in the wild,

Paul Lushenko, Rob Sparrow, and Adam Henschke AI, Trust, Culture, and the Military

in a nonconflict scenario, the failure may not be so bad or so dangerous. So, you have to learn from that failure and then develop and improve the system as a result. But if you're in a conflict environment, the failure there can be significantly catastrophic for people in the field of conflict but also success in warfare.

So, the dangers of failure, or at least partial system failure, are really, really significant here. The hard thing here is how do you test AI in a way that will anticipate and respond to failure such that those failures aren't catastrophic either for the short-term war effort or the overall success of the military campaign?

Paul's typology of the future of AI, for me, I'm really interested in the mosaic human-oversight model that he was talking about. That, to me, sounds both really interesting but also very challenging technically—but also socially or maybe ethnographically. It is both not under trusted nor over trusted. The cultural change there, particularly the work that Paul and Rob have been doing here on the generational issues, are really interesting when you put that in relation to the mosaic stuff.

The final thing that Paul had mentioned about the AI-general sci-fi scenario, that sounds really very, very beneficial in terms of the amount of data that can be assessed. There's possible ways that assessment can be guided in advance through the development of algorithms, et cetera. But it also really lends itself to very obvious sci-fi and nightmare scenarios like your Terminator, Skynet thing.

And obviously, it's always dangerous to go, "Hey, Terminator AI, that's why we shouldn't use AI." But we do need to keep in mind the potential benefits of this sort of AI general but be very wary of the risk. And picking up from what Rob was saying about the flash crash and things like that, where this becomes really complicated and very hard to predict is, you've got systems of systems interacting with other systems of systems, again in a competitive environment, and it can be very, very, very hard to anticipate, predict, or even guess at what those outcomes could be.

The others said it's very hard to predict what will happen here. But those are the sorts of areas that I think are interesting, and things that we need to be thinking a lot about as we move forward in the introduction, or for the development, of these AI systems of systems.

Lushenko

Steph, before we move on, I'd be remiss if I didn't state, given my role here at the US Army War College, that what I attempted to communicate to senior officials that are debating the merits of AI, again, at scale, across domain in the Joint force, are in fact these issues. My big concern is that we have put the cart before the horse without the empirical data to support, and god forbid we actually are in a position where we get that empirical data [about] the usefulness of AI, because I agree a lot with what Rob and Adam said, that there is a strong potential that these capabilities could fail. In fact, Jon Lindsay at Georgia Tech University (Georgia Institute of Technology) will tell us that we should be prepared to be disappointed by AI.

And the other point, too, is that Adam's talked about other questions within this burgeoning research agenda that we haven't focused on as much that bear directly on US military competitive advantage, comparative advantage against our pacing threat, which is China, and that is our allies and partners. It's unclear to me if we could get to the point where we could design AI-enabled systems, however defined, interoperable based upon the systems-to-systems questions that Rob and Adam have both raised.

And so, as I look at the literature on AI, one of the outstanding questions is the implications of culture. How do cultural attributes shape the way that we design AI systems? How does it shape the way we integrate these into our military arsenals? And then furthermore, what's the implication for public attitudes, especially legitimacy, which is really integral to the sustainable use of any modality of force abroad.

This research agenda is so ripe, and it's heartening to see that we have a transnational collection of authors and scholars, at least here in this podcast, that are raising these questions, that I think, for a normative claim, our senior military officials in the United States and Australia, given who we are and where we're coming from, should really pay heed to.

Paul Lushenko, Rob Sparrow, and Adam Henschke AI, Trust, Culture, and the Military

Host

Let's talk a little bit more about trust and the intergenerational challenges here. It seems like younger generations will be more inclined, maybe, to follow orders from an AI commander, or just assume that statements made by AI have some command authority. Will this have implications for the sorts of behaviors seen in future military practice? Are AI trust attitudes likely to lead to increased or decreased problematic behaviors? What are your thoughts?

Sparrow

In a way, I think this question of over trust and under trust that we were talking about previously is the main issue there. If you've grown up with these things gaming [and] if you've used ChatGPT to write your essays or answer your email, perhaps you will be more inclined to trust them when you shouldn't. Because when lives are at stake, as they often are in military settings, those kind of outlier cases, where the systems that are usually quite good but in this case have got something wrong, they become really important.

I'm really interested in an ethical question here, which is, it's one thing to trust AI in relation to decisions that affect the lives of the enemy, or maybe even the lives of noncombatants in the area of operations. In the debate about autonomous weapon systems, there's a big part of the literature that says, "Look, we're gonna trust these systems to make targeting decisions. They're reliable enough. We can get them so that they can do distinction, and they can only attack the right targets."

Well, what if we flip that and the question is, will we trust those systems to make decisions about the lives of friendly forces? When the AI decision-support system tells you, "Look, you should deploy troops five (kilometers or) K down that road," and it's the lives of your comrades at stake, are you gonna trust that? And is it appropriate for commanders to be outsourcing those kinds of decisions to a computer?

But of course, if you're not willing to put friendly lives at stake, there's a real question about noncombatant lives because those lives count here as well. There are these cultural questions, and there are also ethical questions. And then there's the generational stuff. So, there's a lot going on here.

It's pretty easy to be impressed by how quickly the thumbs of your kids move on their mobile phone—younger people are using these systems with a fair amount of confidence. If you push on that a little bit, people are clueless about how the systems work. Their cybersecurity practices are not nearly as good as those of older people. Just because you've grown up with these systems doesn't mean that you actually understand how they work. And [it] doesn't necessarily mean that you're better placed than people who are a bit clunkier, but perhaps rightly suspicious of some of the stuff that's going on.

And so, that's why I think it is important to sort of pay attention to these generational questions. Being more trusting doesn't necessarily correlate with being more competent, for instance.

Henschke

There's a potential issue here with the generational aspect and the increased likelihood to follow commands. Militaries rely on people to follow commands, you know? That's one of the central aspects and defining features, arguably, of what makes [the] military a different institution to a nonmilitary institution.

What we have seen in the past, I don't know, 50, 100 or so years in military training is trying to develop in service members the skills and tools so that they know when to follow commands and when not to follow commands. Now, this is the basic idea of teaching just war at a military college, teaching laws of armed conflict and so on, so that the people who are following commands know, "All right, this is perfectly fine. This is a good command for me to follow, but these are the sorts of commands that I should be questioning, and these are the sorts of commands I shouldn't follow."

If there is this increased generational trend to be more trusting of AI, we need to develop a set of practices, and also a set of design features, around our AI and integration of AI into military practice—such that people have a similar

Paul Lushenko, Rob Sparrow, and Adam Henschke AI, Trust, Culture, and the Military

capacity to judge when it is better to be pro trust, and when it is better to be less trusting. And I think that can take lessons from, again, the education of just war, the education of laws of armed conflict.

Obviously, [this is a] similar context, but it would need to be conducted in quite a different way. But to build it into the militaries—and particularly for these younger generations, the skills and tools—so that we can better calibrate the over trust and under trust issues, I think that’s one way of recognizing this issue and trying to respond to it, in advance, and hopefully, in a bit of a better way than we might have done in the past.

Lushenko

I think that’s right. And in other related research, I take a first stab at what factors—when brought together, in a vacuum, at least, in a sterile, artificial world—could, in fact, increase/decrease trust in partnering with these systems. And what I find—at least among senior military officers at the O-5 lieutenant colonel and higher, across both the Army and the Navy—is that this is really a multidimensional and complex problem set.

And it’s a function of technical aspects to include the autonomy level, the degree of precision, the type of supervisory oversight. It’s a function of, frankly, countervailing moral logics, in terms of what Rob talked about, force protection, but also buying down the risk to civilians and collateral damage, what we call, in terms of law of war, distinction and proportionality.

And then, finally, is, in terms of these operational considerations, the likelihood of success for the use of these capabilities, for overall mission accomplishment. Another consideration would be regulatory oversight, whether it’s domestically levied or internationally levied. And I think the real challenge, although I agree with Adam, is that we can, in this research, identify these factors, [and] when brought together, [they] can increase levels of trust. But what does that look like in a hot war between near-peer competitors? To what extent will we rely upon these moral precepts, in an existential environment where the threat, potentially, is against our sort of way of life?

And I think this is really important because one of the things we just talked about today at the (US) Army War College, through our oral comprehensive exams, are the normative dimensions of the use of force that underpin considerations on force employment. The legal considerations that Adam talked about [and] the moral, and the ethical.

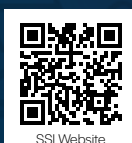
Host

Don’t miss episode two, coming soon.

Download this episode:

<https://ssi.armywarcollege.edu/SSI-Media/Podcasts-Lectures-and-Panels/Conversations-on-Strategy-Podcast/mod/67381/player/581/audio/80480>

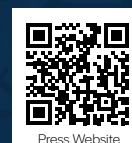
More information about the programs of the Strategic Studies Institute (SSI) and US Army War College (USAWC) Press can be found on the Institute’s home web page at <https://ssi.armywarcollege.edu/>.



SSI Website



This Episode



Press Website