By *Dis*analogy, Cyberwarfare is Utterly New

Selmer Bringsjord
Rensselaer Polytechnic Institute
Troy, NY, USA
selmer@rpi.edu

John Licato
Rensselaer Polytechnic Institute
Troy, NY, USA
licatj@rpi.edu

Abstract:
We provide an underlying theory of argument by *dis*analogy, in order to employ it to show that cyberwarfare is fundamentally new (relative to traditional kinetic warfare, and espionage). Once this general case is made, the battle is won: we are well on our way to establishing our main thesis: that Just War Theory itself must be modernized. Augustine and Aquinas (and their predecessors) had a stunningly long run, but today's world, based as it is on digital information and increasingly intelligent information-processing, points the way to a beast so big and so radically different, that the core of this duo's insights needs to be radically extended.

Keywords: Cyberwarfare, Analogy, Disanalogy

1. Introduction

The reader is likely familiar with the claim that cyberwarfare is fundamentally nothing new. While it may be that some proponent of this claim will arrive on the scene touting a purely deductive argument for it from axioms that self-evidently capture all forms of warfare, this seems rather unlikely. We do have axioms for such things as number theory, from which at least the vast majority of arithmetic can be deductively derived. But war is different; very different. War is hell, yes; but it is also something that subsumes every aspect of the not-yet-even-remotely-formalized world of human cognition, perception, action, and emotion — *and* involves and requires intimate command of the mechanico-physical world of weapons and their effects. As powerful as the traditional axiomatic method may be, in the face of the vast, towering complexity of the target here, that method, at least in its standard form, appears anemic. Attempts to establish the proposition that cyberwarfare is old hat must find their foundation not in mere deduction, but something else, at least primarily. But what?

The answer seems clear, actually: Those advancing the claim that cyberwarfare is just a wrinkle at the level of details far beneath the *nature* of warfare and the ethics thereof, must rely, crucially, on *analogical* reasoning; that is, the core idea must be that cyberwarfare can be shown by analogy to at its heart be no different than longstanding X. For instance, it is fair to say that before the advent of emails and hyperlinks within them, spear phishing didn't exist; yet today, if Jones receives an email that fits perfectly within the context of life working under and for his superior, and which asks him to click here to receive the latest draft of the report the team is working on, he may well do so — even if the email is from the enemy. If we let X be espionage, then the analogical argument in the case now at hand, in short, is that while this sort of thing is *specifically* new, it's really just analogous to any number of ruses perpetrated by clever spies from time immemorial. Spies have long been forging documents, after all; and an email with a hotlink is — so the story goes — no different, really, than a forged hard-copy document with a request in it. If Just War Theory (JWT) provides verdicts with respect to familiar forgery and ruse, then, so the story continues, it must provide a verdict in the case of spear phishing. A similar analogical story could be told in connection with cyberphysical attacks: If the tank that Smith is driving can be disabled by a remote enemy hacker who compromises the "shroud" of software that, increasingly, high-tech vehicles are cradled in, well, that is significant; but why — so another such story goes — is such an attack fundamentally different than an enemy soldier blasting the tank with a kinetic weapon from close range?

We believe that cyberwarfare (along with some forms of "mild" cyberconflict) is not only fundamentally new, but, upon closer inspection, *dangerously* new. In order to defend our position, we could do merely what we find some likeminded colleagues doing, and what we have ourselves been tempted to do: rebut analogical arguments on a case-by-case basis, over and over, showing in each case that the presumed analogy doesn't in fact hold. Such individual-case refutations would of course employ a theory of analogical argumentation (such as Bartha 2010), and show that the normative structure of such argumentation, according to the theory, isn't fully satisfied in the particular case in question. But this is surely a most inefficient approach, and is probably a losing battle for anyone who, like us, sees cyberwarfare, now only in its infancy, to in the future be of paramount importance.

Instead, we shift from the defensive to the offensive mode. We analyze traditional warfare, espionage, and cyberwarfare, producing not an axiomatic system for each poised to serve as a source of deduction, but producing instead a representative quartet of necessary conditions (for each concept) sufficient to undergird rigorous

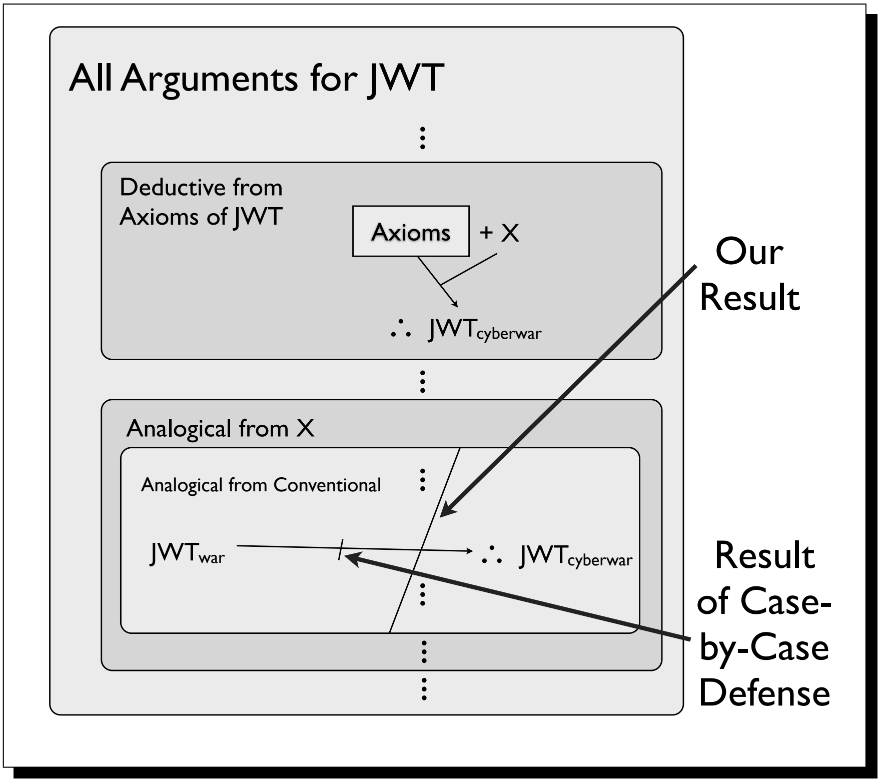


Figure 1: The argument we outline in this paper aims to show not that some *particular* argument by analogy fails to hold between these two domains, but rather that these domains are sufficiently dissimilar that they cannot possibly produce *any* good analogy.

disanalogical reasoning. Next, via (deductive) meta-reasoning (over an instantiation of an underlying theory of argument) that shows *dis*analogy, we show that cyberwarfare is fundamentally new. While cognitive science and artificial intelligence have seen much fruitful effort devoted to sorting out argumentation that appeals to analogies, no one has sorted out the technique of systematically finding and exploiting, in argument, *dis*analogy.[1] Once our general case is made, the battle is won: we are well on our way to establishing our main thesis: that JWT itself must be modernized. Augustine and Aquinas[2] had a stunningly long run, but today's world, based as it is

[1] There is work that suggests analogical mechanisms are at play in determining which features are most salient in similarity and dissimilarity assessments (Markman & Gentner 1993, Gentner & Sagi 2006). However, this work is descriptive, rather than normative.

[2] And their likeminded predecessors, of course. E.g., cleary Cicero deserves credit for articulating aspects of modern JWT. But the present paper is premeditatedly not heavy on scholarship, since its focus is instead on the inductive logic of traditional reasoning offered in support of the "nothing new" view of

on digital information and increasingly intelligent information-processing, points the way to a beast so big and so radically different, that the core of this duo's insights need to be radically extended.[3]

The plan for the sequel is as follows. We next (§2) give a brief review of prior work on analogical reasoning carried out in our laboratory. We then (§3) set out a general schema for analogical argumentation that any worthwhile analogical argument must abide by. Next, in section 4, we describe the instantiation of this schema in which an inference is made from the applicability of JWT in the conventional case, to the proposition that JWT applies to the real of cyberwarfare. The next section (5) is devoted to showing that because essential attributes of the SOURCE domain (conventional warfare and espionage) are lacking in the TARGET domain, the entire space of analogical arguments from the applicability of JWT in the conventional sphere, to the applicability of JWT in cyberwarfare, is defective. (Our argument, placed in contrast to case-by-case arguments, can be visualized as in Figure 1.) Some concluding remarks close the paper.

2. Prior Work on Analogical Reasoning

The topic of this paper is smoothly and firmly in line with a general direction the RAIR Lab has been pursuing: namely, the intersection of logic, analogy, and AI. One specific piece of this prior work has been devoted to what we have coined **analogico-deductive reasoning** (ADR): the combination of analogical and hypothetico-deductive reasoning, as described for instance in Licato and Bringsjord (2012). In ADR, a common reasoning process used by children when solving Piagetian puzzles (Bringsjord & Licato 2012) (see Figure 2) all the way to master mathematicians and logicians establishing profound theorems (Licato, Govindarajulu, Bringsjord, Pomeranz & Gittelson 2013, Licato, Bringsjord & Govindarajulu 2013), analogy is used to generate a hypothesis *h* about some target domain. Deductive reasoning is then used to either support or falsify *h*. For example, Licato et al. (2013) demonstrated an ADR system that took as input the proof of the so-called "Liar Paradox," some axioms from mathematical logic, and some domain knowledge. The system was able to draw an analogy from the proof of the Liar Paradox to a proof of Gödel's First Incompleteness Theorem (**G1**), and fill in the gaps of the proof, resulting in the high-level proof pictured in Figure 3.

Our work on ADR has hitherto focused on the cognitive dimension of this form of reasoning: identifying its use, modeling it, and computationally simulating that use. This paper shifts focus to the *argumentation* side of things, in connection with a pressing issue of the day. We know that ADR is common in human reasoning, and on the strength of our research to this point, we know as well that ADR can be rendered rigorous and implemented; but our results say little about judging whether real humans tackling real issues using analogical argumentation are reasoning correctly. In ADR, the deductive component helps to ensure that, given the state of the axioms from which the deduction is derived, conclusions have some guarantee of correctness. However, analogical reasoning carried out by flesh-and-blood humans in high-stakes domains is often entirely separate from deduction, and by its very nature

cyberwarfare. Accordingly, we don't spend time rehearsing the roots of JWT, which are likely to be familiar to all of our readers.

[3] We do in fact believe the core *can* be extended. Nothing we say herein should be taken to impugn the "meta-JWT" ethical core of Augustine and Aquinas. We are not prepared to issue such a vote of confidence in other such cores, from other thinkers.

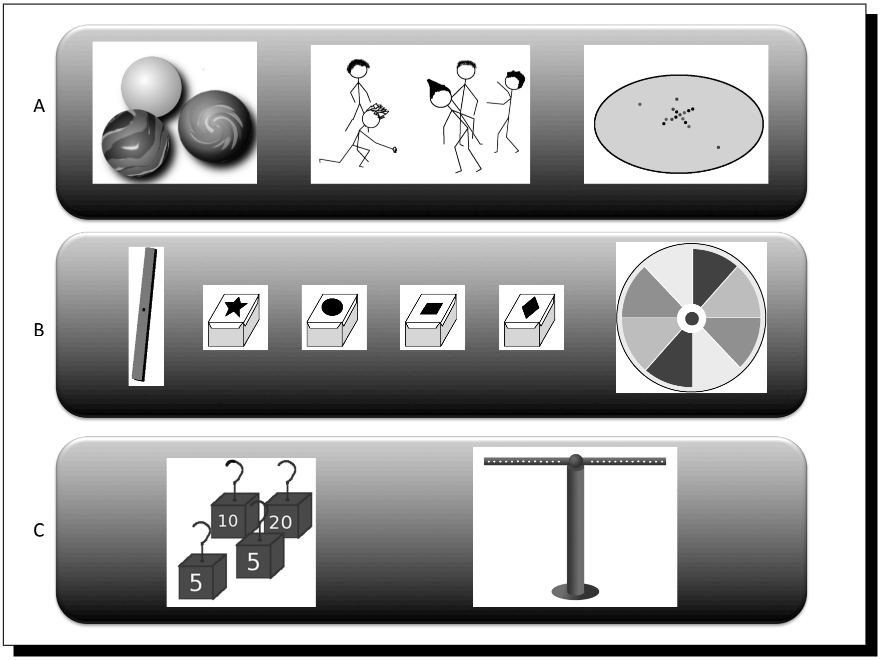
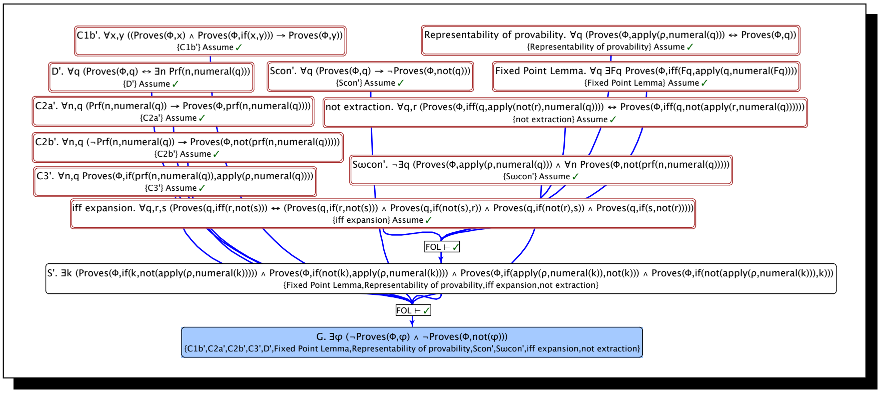


Figure 2: Objects and Concepts for a “Piaget-MacGyver Room” Experiment. See Bringsjord and Licato (2012) for more details.



C1b'. ∀x,y ((Proves(Φ,x) ∧ Proves(Φ,if(x,y))) → Proves(Φ,y))
{C1b'} Assume ✓

Representability of provability. ∀q (Proves(Φ,apply(ρ,numeral(q))) ↔ Proves(Φ,q))
{Representability of provability} Assume ✓

D'. ∀q (Proves(Φ,q) ↔ ∃n Prf(n,numeral(q)))
{D'} Assume ✓

Scon'. ∀q (Proves(Φ,q) → ¬Proves(Φ,not(q)))
{Scon'} Assume ✓

Fixed Point Lemma. ∀q ∃Fq Proves(Φ,iff(Fq,apply(q,numeral(Fq))))
{Fixed Point Lemma} Assume ✓

C2a'. ∀n,q (Prf(n,numeral(q)) → Proves(Φ,prf(n,numeral(q))))
{C2a'} Assume ✓

not extraction. ∀q,r (Proves(Φ,iff(q,apply(not(r),numeral(q)))) ↔ Proves(Φ,iff(q,not(apply(r,numeral(q))))))
{not extraction} Assume ✓

C2b'. ∀n,q (¬Prf(n,numeral(q)) → Proves(Φ,not(prf(n,numeral(q)))))
{C2b'} Assume ✓

Sωcon'. ¬∃q (Proves(Φ,apply(ρ,numeral(q))) ∧ ∀n Proves(Φ,not(prf(n,numeral(q)))))
{Sωcon'} Assume ✓

C3'. ∀n,q Proves(Φ,if(prf(n,numeral(q)),apply(ρ,numeral(q))))
{C3'} Assume ✓

iff expansion. ∀q,r,s (Proves(q,iff(r,not(s))) ↔ (Proves(q,if(r,not(s))) ∧ Proves(q,if(not(s),r)) ∧ Proves(q,if(not(r),s)) ∧ Proves(q,if(s,not(r)))))
{iff expansion} Assume ✓

FOL ⊢ ✓

S'. ∃k (Proves(Φ,if(k,not(apply(ρ,numeral(k))))) ∧ Proves(Φ,if(not(k),apply(ρ,numeral(k)))) ∧ Proves(Φ,if(apply(ρ,numeral(k)),not(k))) ∧ Proves(Φ,if(not(apply(ρ,numeral(k))),k)))
{Fixed Point Lemma,Representability of provability,iff expansion,not extraction}

FOL ⊢ ✓

G. ∃φ (¬Proves(Φ,φ) ∧ ¬Proves(Φ,not(φ)))
{C1b',C2a',C2b',C3',D',Fixed Point Lemma,Representability of provability,Scon',Sωcon',iff expansion,not extraction}

Figure 3: Full Deductive “Short-Distance” Proof of **G1** in Slate, Automatically Generated. See Licato et al. (2013) for more details.

is error-prone. A flawed commonsense understanding of analogy, and specifically of how and when analogical argumentation works best, has bred many false analogies and dismissal of productive analogies. A lack of attention to this shortcoming is especially harmful when dealing with phenomena having two general attributes: phenomena that are at once new and unfamiliar, since in such domains analogies are often a centrally important tool for understanding concepts; and secondly, phenomena such that, when mistakenly analyzed, can have serious real-world consequences. Cyberwarfare meets both of these conditions.

3. Generic, Unexceptionable Argument Schema

This section is devoted to setting out a generic, unexceptionable schema $\mathscr{A}$ for analogical argumentation.

For ease of exposition, assume a domain of discourse for both the SOURCE and TARGET (D and D^*, resp.), and suppose as well that there is a set of sets of formulae in some language for each of both the SOURCE and TARGET ($\mathcal{L}_1, \mathcal{L}_2$, resp.); these formulae express information about the SOURCE and TARGET, and contain specifically all the relation symbols and function symbols needed to make relevant assertions, including — as JWT requires — assertions about what ought to be done and what is forbidden.[4] Note that there is a key particular formula χ that holds of the SOURCE, whose analogue, χ^*, is the specific thing inferred to hold about the TARGET. The situation is shown schematically in tabular form in Table 1. This table indicates that an analogical mapping holds between the set

$$\mathcal{P} = \{\phi_1, \phi_2, \ldots, \phi_n\}$$

and

$$\mathcal{P}^* = \{\phi_1^*, \phi_2^*, \ldots, \phi_n^*\}.$$

As is well-known, there are helpful positive mappings that hold between the domain of water flow and the domain of electricity. Though this is very crude, in keeping with the positive mapping here from $\mathcal{P}$ to $\mathcal{P}^*$, it may help to imagine that the former set includes a formula $\phi_i(F, P)$ in which F and P are predicate letters representing the ordinary, intuitive attributes *Flows* and *Pipe*, respectively, which range over the domain of water and its movement. (We are not concerned with what $\phi_i(F, P)$ specifically says about water flow; it suffices to have in mind that this formula in general asserts that water in a plumbing system flows through pipes.) Now imagine in addition that the domain is shifted to the TARGET: electricity. The corresponding formula $\phi_i^*(F^*, P^*)$ now says that electricity "flows" through "pipes" (= wires). This mapping, as a matter of fact, is often used to explain electricity to those unfamiliar with it, but familiar with the basic plumbing concepts (Gentner & Gentner 1983).

In addition, there is the negative part of the schema. This is indicated by the fact that while the formulae in $\mathcal{A}$ hold of SOURCE, they don't hold of TARGET; and by the fact that while the formulae in $\mathcal{B}$ fail to hold of SOURCE, they do apply to TARGET. Finally, the reader will notice a line in the schema that serves as a placeholder ready to receive any number of proposed conditions that must be satisfied in order for

[4] Clearly, both $\mathcal{L}_1 \mathcal{L}_2$ will need to be formal languages that each include formal sub-languages for robust deontic logic. Deontic logics are deployed to formalize ethical principles. For the classic introduction to deontic logic in just a few elegant pages, see (Chellas 1980). For a non-technical discussion of the use of computational deontic logic to govern artificial intelligents, see (Bringsjord, Arkoudas & Bello 2006). In addition, these languages will need to subsume languages that allow for modeling of belief, desires, intention, perception, and communication. For a computational logic in exactly this direction, see (Arkoudas & Bringsjord 2009).

Table 1: Generic Schema $\mathscr{A}$ of Analogical Argument

SOURCE		TARGET
$\mathcal{P}$	$\longrightarrow$	$\mathcal{P}^*$
$\mathcal{A}$	$\not\longrightarrow$	$\neg\mathcal{A}^*$
$\neg\mathcal{B}$	$\not\longrightarrow$	$\mathcal{B}^*$
proposals		*proposals*
χ		
		χ^*

the inference to go through. For example, it has been suggested that the formulae in $\mathcal{P}$ must be "relevant" to χ, and so on. Because of the nature of our reasoning herein, that is, because our focus is on deducing *dis*analogy, we have no need to explore these additional conditions, and are left to sedulous readers to investigate.[5]

It's crucial to understand that the schema $\mathscr{A}$ given here is unexceptionable. That is, any successful analogical argument for an ultimate conclusion χ^* will be an instance of this schema. The schema puts no one offering an analogical argument at a disadvantage, and simply reflects the underlying, immovable formal reality behind any analogical argument. The broad applicability of $\mathscr{A}$ is of course a key part of our recipe, which, recall, is to show that an entire range of instantiations of the scheme, namely those that purport to establish the applicability of JWT to cyberwarfare, are fatally flawed (see again the larger stroke '/' in Figure 1). We turn now to a characterization of the range in question.

4. The Instantiated Generic Argument Schema

This section is devoted to presenting the instantiation of the generic schema $\mathscr{A}$ for analogical argumentation to the purpose of showing that cyberwarfare, from the standpoint of JWT, is nothing new.

We assume a domain of discourse for both the SOURCE and the TARGET (D_{war} and D^*_{cyberwar}, resp.), and suppose as well that there is a set of sets of formulae (in $\mathcal{L}^1_{\text{war}}$ and $\mathcal{L}^2_{\text{cyberwar}}$, resp.) for each of both the SOURCE and TARGET; these formulae express information, of course, regarding the SOURCE and TARGET. In addition, there is a key particular formula $\mathbf{JWT}_{\text{war}}$ that holds of the SOURCE, whose analogue, $\mathbf{JWT}_{\text{cyberwar}}$, is inferred to hold about the target. The formula $\mathbf{JWT}_{\text{war}}$ represents the overall claim that JWT applies to cyberwarfare, in a direct "carry over" from its application to conventional war and espionage. The situation is shown schematically in tabular form in Table 1.

[5] The place to start is the *Stanford Encyclopedia of Philosophy* entry "Analogy and Analogical Reasoning" by Paul Bartha: He considers a number of possibilities for filling in *proposals* in $\mathscr{A}$, a schema which, in agreement with our position herein, he endorses as an unobjectionable starting place for capturing the structure of analogical argumentation in science. See (Bartha 2013).

Table 2: Instantiation $\mathscr{A}^{w \to c}$ of Generic Schema of Analogical Argument

SOURCE		TARGET
$\mathcal{P}_{\text{war}}$	$\longrightarrow$	$\mathcal{P}^*_{\text{cyberwar}}$
$\mathcal{A}_{\text{war}}$	$\not\longrightarrow$	$\neg\mathcal{A}^*_{\text{cyberwar}}$
$\neg\mathcal{B}_{\text{war}}$	$\not\longrightarrow$	$\mathcal{B}^*_{\text{cyberwar}}$
proposals		*proposals*
$\mathbf{JWT}_{\text{war}}$		
		$\mathbf{JWT}_{\text{cyberwar}}$

5. Negation of Essential Attributes in SOURCE

We have let $\mathbf{JWT}_{\text{war}}$ denote the collection of ethical principles ranging over humans and the elements of their relevant cognition (knowledge, belief, actions, intentions, etc.), weapons, psychological techniques, ruses, and so on. And we have denoted the collection of principles that are to be analogically transferred to the realm of cyberwarfare by $\mathbf{JWT}_{\text{cyberwar}}$, in accordance with the argument schema $\mathscr{A}^{w \to c}$ (which is of course generates a proper subset of arguments defined by $\mathscr{A}$) set out above. In this schema, $\mathcal{A}_{\text{war}}$ refers to truths about the conventional SOURCE that fail to hold with respect to the TARGET. (For example, pulling the trigger of conventional, purely kinetic gun in a firefight is a kind of action found in the SOURCE, but not in the TARGET.) But we also know that when the propositions in $\mathcal{A}_{\text{war}}$ include *essential* aspects of the SOURCE, any analogical argument that includes this "mismatch" is vitiated. But there are indeed certain essential truths about the SOURCE that fail to transfer to the TARGET; in fact, there are *many* such discrepancies. We show this in the next section, but first pause to give what we hope is an illustrative example.

Suppose that some thinker wants to argue for the — radical, yes — proposition ($\bar{E}^*$) that human beings, even adult neurobiologically normal and well-nurtured ones, aren't bound by any ethical prohibitions whatsoever. This thinker gives an analogical argument for $\bar{E}^*$, the corresponding SOURCE proposition for which is encoded in the TARGET as $\bar{E}$. Our hypothetical thinker's main move, we imagine, is that $\bar{E}^*$ follows from the fact that "mere" animals (e.g., mice, cats, dogs, etc.) don't have the cognitive capacity to understand ethical principles, and their bases. To make this move a bit more concrete, suppose for the sake of argument that some action a is obligatory for an agent if and only if a, above all competitors at the relevant time, secures consequences that have the greatest utility. In this consequentialist context, our thinker points out that dogs surely lack the intellectual power to understand and apply this principle; hence the thinker cheerfully that canines aren't morally obligated to perform any actions. Since, as our thinker explains, parallel reasoning holds for mere animal after mere animal, $\bar{E}$ holds.

Next, our thinker proceeds to flesh out his instance of the schema $\mathscr{A}$ by pointing to a number of mappings from propositions regarding mere animals, to propositions about humans. For example, he points to similarities between the genetic material of mere animals and the genetic material of *homo sapiens sapiens*, similarities at the level of physiology and anatomy, and so on. This is to say that he here fleshes out his particular instance of the mapping from $\mathcal{P}$ to $\mathcal{P}^*$. Our thinker also points out that mere animals are conscious, as are humans. The story here could be expanded greatly (we could for instance restrict it to mammals on the animal side, and we could point to many additional similarities). In addition, and finally, we can suppose that our thinker does instantiate the negative side of $\mathscr{A}$ — but in a tendentious manner. For example, he concedes that many animals are not bipedal, whereas humans are.

It will of course be obvious to the reader that our thinker's analogical case for $\bar{E}^*$ is doomed. Why? The fatal flaw is that *essential* attributes on the SOURCE side are in the negative portion of $\mathscr{A}$. Yet as Hesse (1966) has convincingly argued, perhaps the chief requirement for an analogical argument to establish (or at a minimum lend credence to) some conclusion χ^* is that the negative side (i.e., the mismatching $\mathcal{A}$ and $\mathcal{B}$ in $\mathscr{A}$) *not* include essential attributes. In the parable under consideration, the problem is of course specifically that it's part of the very nature of being a *mere* animal that there is sub-human capacity cognitively and linguistically. Put another way, mere animals aren't *persons*, yet it is the qualities that constitute personhood that hold sway in matters moral.[6] In our disproof, likewise, we show that essential attributes of the SOURCE are indeed absent in the TARGET. What are the attributes in question? We turn to them now.

5.1. Essential Attributes of the SOURCE

In order to advance our case clearly despite space constraints, we focus our attention within JWT on *jus in bello*, and further focus our attention upon four uncontroversial attributes that are essential to the SOURCE in connection with *jus in bello*.[7]

- *Control.* Weapons that aren't controllable are, under JWT, immoral to deploy. This is presumably why certain biological weapons are immoral. For instance, use of a mysterious but deadly and highly contagious biological virus would be prohibited under *jus in bello*, in significant part because to unleash this weapon would, for all the user knows, result in harm that is unimaginably severe, and that is entirely chaotic. It is thus essential to the propositions characterizing the SOURCE in $\mathscr{A}^{w\rightarrow c}$ that the effects of conventional weapons and techniques be generally assessible to human cognition.
- *Proportionality.* This familiar set of principles dictates that in just war war no attacks can be disproportional to the ends sought. It is essential to these principles that warfighters have an understanding about the effects of the actions that they can perform, since without that understanding there would be no way to form in the first place a rational belief about what is proportional and what isn't.
- *Discrete, Directly Accessible Analog Objects.* Here we refer to a set of truths about conventional warfare and espionage that are presupposed by *jus in bello*; namely, that warfighters can perceive ordinary physical objects and their boundaries, that they can access (e.g., manipulate) these objects, that these objects travel in standard spatiotemporal arcs, and so on.
- *Discrimination and Non-Combatantant Immunity.* Here there are of course obligations in force that require warfighters distinguish between innocent non-combatants versus

[6] For a discussion of personhood (including a definition thereof) as the crux of a reasoned discussion of a profound ethical issue, see (Bringsjord 1997).

[7] More formally, there will be formulae in $\mathcal{A}_{\text{war}}$ that make use of the predicate symbols and function symbols used to express the quartet of attributes enumerated by us here.

> combatants, and that they refrain from intentionally harming persons in the former category in the course of seeking military victory. The obligations in question undeniably presuppose not only that discrimination can in fact take place, but that actions can be selected on the basis of whether or not they impact non-combatants. Specifically, warfighters are here assumed to be able to carry out courses of action that impact combatants, but not non-combatants (at least not directly).

We assume that it's clear how this quartet implies a host of attributes that necessarily hold of objects in the SOURCE. For example, we can say, on the strength of the simple inventory of *jus in bello* just taken, that, necessarily, if *a* is a human warfighter, then he or she is able to steer clear of actions in conflict that may very well propagate across the globe indiscriminately, harming combatants and non-combatants alike, in all manner of nation or group.

We turn now to a brief discussion of the future of AI, the field devoted to building intelligent agents, including autonomous ones (Russell & Norvig 2009).

5.2. The MiniMaxularity, Cyber, and Our Future

Many readers will be familiar with The Singularity, that future moment when machines with human-level intelligence move beyond that level, and then exploit their superhuman powers to built smarter and smarter and ... smarter machines, leaving us in the dust. The main argument for the proposition that The Singuarity will occur is first given by Good (1965), and is ably amplified by Chalmers (2010) — but the argument, which is by the way not an analogical one, but a deductive one, need not concern us here. Under not-unreasonable mathematical assumptions, one of us has proved that The Singularity is impossible (see Bringsjord 2012); but this purported refutation can also be left aside, given present purposes. What *is* relevant to the present paper is the concept of The MiniMaxularity, introduced by S. Bringsjord and A. Bringsjord in the forthcoming paper "The Singularity Business."[8] This is the concept that machine intelligence will indeed reach great heights, but will be "*mini*mal" relative to The Singularity (e.g., machines will not have subjective awareness or self-consciousness), yet "*max*imal" with respect to certain logico-mathematical constraints.[9] These constraints, put rather impressionistically here, which is sufficient for present objectives, amount to saying that computing machines will reach a level of intelligence that is maximal along the lines of the smartest such machines we have so far seen. A paradigmatic example of such a machine is IBM's Watson, a QA system that famously defeated the two best *Jeopardy!* players on our planet. Watson was engineered in short order, by a tiny (but brilliant and brilliantly led) team, at a tiny cost relative to the combined size of today's AI companies, which includes Google, at its heart certainly an AI company.[10] We assume that The MiniMaxularity will occur, and our case for the novelty of cyberwarfare is couched in terms of its arrival in the future. Note that the combined market capitalization of just the large AI companies of today is probably over one trillion (U.S.) dollars. Given those resources, imagine what will be in place, say, 20 years from now with respect to intelligent agents. At that time, barring some global catastrophe (e.g., Earth gets hit by an asteroid), it will be possible to ask AIs to go into cyberspace and try to do

[8] Which is in turn forthcoming in the book *The Technological Singularity: A Pragmatic Perspective*. For the position that The Singularity is a pipe-dream, see (Bringsjord, Bringsjord & Bello 2013).

[9] This is basically the vision explicated in (Bringsjord 1992).

[10] For an overview of Watson, see (Ferrucci et al. 2010).

any number of nasty things — and these AIs will be autonomous and powerful to the point that ordinary human minds will have precious little understanding of how these AIs work, and what, unleashed to their own devices, they will do.

In addition to assuming The MiniMaxularity, we assume that at least the vast majority of the ordinary analog world will, in concert with the arrival of smart and autonmous AIs, be completely enveloped or enshrouded in digital software. (We have here been heavily influenced by Luciano Floridi, who has an uncanny ability to see the future in connection with information.) There will be no such thing, in our future, as a physical weapon in the ordinary sense. If today you handed Augustine a standard kinetic gun, he wouldn't have much trouble grasping (given an explanation, one inevitably based on drawing analogies to the weapons of his time) the nature of what you had given him. But in the future, it will not be possible to access a gun *qua* gun. Instead, the physical will be buried under complex cyber layers constituted by software. Indeed, we believe that it will be impossible to access ordinary kinetic weapons without first engaging AIs that are inextricably bound up in these cyber layers. They layers of software that will enshroud all things analog will be shot through and through with agents that have never been part of warfare.

5.3. The Disproof

Given how we have set the table with the preceding content and discussion, an outright disproof of the claim that JWT applies to cyberwarfare be easy to obtain, and we give here only the informal proof-sketch: We first simply note that in the future, intelligent autonomous agents will be part of the digital bloodstream of our planet, and that that bloodstream will enshroud standard kinetic causation within a digital world, so that there for instance be no such thing as "pulling a trigger." Indeed, and in short, *all* of the essential attributes called out in our enumeration of the quartet above (§5.1) fail to hold in the realm of cyberwarfare as we depict it. For example, releasing an AI with the task of disabling a nuclear arsenal by disabling the software shroud around that arsenal may for all anyone knows unleash destructive forces that are disproportional and which greatly impact non-combatants. We next simply note that it's a necessary truth that conventional warfare and espionage satisfy the conditions enumerated in section 5.1. Given the aforementioned principle that any analogical argument in which the SOURCE's essential attributes are in the negative side of $\mathscr{A}$, we deduce the result that every analogical argument within the class of arguments relied upon by proponents of the "cyber is nothing new" view fails. QED. Of course, this is not to say that there isn't *another* route of reasoning for such proponents to try to find. We here only close off one type of route to the "nothing new" position, and close off thereby the applicability of JWT to cyberwarfare that would be entailed by that position.

6. CONCLUSION

We have presented a deductive case for the proposition that cyberwarfare is fundamentally new, and that therefore Just War Theory, long readily and indeed comfortably applied to conventional warfare and espionage, does *not* apply to cyberwarfare. We gladly concede that our case at this early point in its evolution is not only inchoate, expressed as it is within but a few short pages, but also concede that our argument has premises that are far from self-evident. There will doubtless be readers who refuse to accept our prediction that The MiniMaxularity will soon enough arrive, and that hyper-complex computation will entirely cloak every single traditional physical object and event of a type that warfighters from time immemorial have studied and exploited.

As to future work, well, obviously, a prime challenge is to formulate an ethic for cyberwarriors that applies to a future in which AIs of great reach, power, and independence roam everywhere among us, and in which the kinetic currency of war is pushed down to a remote distance far removed from where the real economy of conflict will ebb and flow, moved by the behavior of computer programs.

Acknowledgments

We are indebted to both AFOSR and ONR for sponsoring our research aimed at giving computing machines, including robots, an ability to reason analogically. This research uncovered for us the surprising fact that while many have sought to set out the structure of analogy, and reasoning that exploits analogy, few to none have set their sights on spelling out *dis*analogy, and on devising systematic processes that leverage this phenomenon. We confess that prior work by Bringsjord, based on the idea that even futuristic robots on the battlefield pose no particular problems for the application of conventional ethics, may need to be revised in light of the coming MiniMaxularity and the concomitant ubiquitous cloaking of the "normal" physical world with software. We are grateful to Tamara Schwartz for emphasizing that in the cyberworld, actions aimed at combatants will inevitably effect non-combatants. Finally, many thanks to NATO CCD COE for the opportunity to share our ideas, the future refinement and defense of which is a burden we accept.

REFERENCES

Arkoudas, K. & Bringsjord, S. (2009), 'Propositional Attitudes and Causation', *International Journal of Software and Informatics* **3**(1), 47–65.
URL: *http://kryten.mm.rpi.edu/PRICAI_w_sequentcalc_041709.pdf*

Bartha, P. (2013), Analogy and Analogical Reasoning, *in* E. Zalta, ed., 'The Stanford Encyclopedia of Philosophy'.
URL: *http://plato.stanford.edu/archives/fall2013/entries/reasoning-analogy*

Bartha, P. F. (2010), *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*, Oxford University Press.

Bringsjord, S. (1992), *What Robots Can and Can't Be*, Kluwer, Dordrecht, The Netherlands.

Bringsjord, S. (1997), *Abortion: A Dialogue*, Hackett, Indianapolis, IN.

Bringsjord, S. (2012), 'Belief in The Singularity is Logically Brittle', *Journal of Consciousness Studies* **19**(7), 14–20.
URL: *http://kryten.mm.rpi.edu/SB_singularity_math_final.pdf*

Bringsjord, S., Arkoudas, K. & Bello, P. (2006), 'Toward a General Logicist Methodology for Engineering Ethically Correct Robots', *IEEE Intelligent Systems* **21**(4), 38–44.
URL: *http://kryten.mm.rpi.edu/bringsjord_inference_robot_ethics_preprint.pdf*

Bringsjord, S., Bringsjord, A. & Bello, P. (2013), Belief in the Singularity is Fideistic, *in* A. Eden, J. Moor, J. Søraker & E. Steinhart, eds, 'The Singularity Hypothesis', Springer, New York, NY, pp. 395–408.

Bringsjord, S. & Licato, J. (2012), Psychometric Artificial General Intelligence: The Piaget-MacGyver Room, *in* P. Wang & B. Goertzel, eds, 'Theoretical Foundations of Artificial General Intelligence', Atlantis Press, 8, square des Bouleaux, 75019 Paris, France.
URL: *http://kryten.mm.rpi.edu/Bringsjord_Licato_PAGI_071512.pdf*

Chalmers, D. (2010), 'The Singularity: A Philosophical Analysis', *Journal of Consciousness Studies* **17**, 7–65.

Chellas, B. F. (1980), *Modal Logic: An Introduction*, Cambridge University Press, Cambridge, UK.

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A., Lally, A., Murdock, W., Nyberg, E., Prager, J., Schlaefer, N. & Welty, C. (2010), 'Building Watson: An Overview of the DeepQA Project', *AI Magazine* pp. 59–79.
URL: *http://www.stanford.edu/class/cs124/AIMagzine-DeepQA.pdf*

Gentner, D. & Gentner, D. R. (1983), Flowing Waters or Teeming Crowds: Mental Models of Electricity, *in* D. Gentner & A. Stevens, eds, 'Mental Models', Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 99–129.

Gentner, D. & Sagi, E. (2006), Does "Different" Imply a Difference? A Comparison of Two Tasks, *in* R. Sun & E. Sagi, eds, 'Proceedings of the 28th Annual Conference of the Cognitive Science Society'.

Good, I. J. (1965), Speculations Concerning the First Ultraintelligent Machines, *in* F. Alt & M. Rubinoff, eds, 'Advances in Computing', Vol. 6, Academic Press, New York, NY, pp. 31–38.

Hesse, M. (1966), *Models and Analogies in Science*, University of Notre Dame Press, Notre Dame, IN.

Licato, J., Bringsjord, S. & Govindarajulu, N. S. (2013), How Models of Creativity and Analogy Need to Answer the Tailorability Concern, *in* T. R. Besold, K.-u. Kühnberger, M. Schorlemmer & A. Smaill, eds, 'Proceedings of the IJCAI 2013 Workshop on Computational Creativity, Concept Invention, and General Intelligence', Beijing, China.

Licato, J., Bringsjord, S. & Hummel, J. E. (2012), Exploring the Role of Analogico-Deductive Reasoning in the Balance-Beam Task, *in* 'Rethinking Cognitive Development: Proceedings of the 42nd Annual Meeting of the Jean Piaget Society', Toronto, Canada.
URL: *https://docs.google.com/open?id=0B1S661sacQp6NDJ0YzVXajJMWVU*

Licato, J., Govindarajulu, N. S., Bringsjord, S., Pomeranz, M. & Gittelson, L. (2013), 'Analogico-Deductive Generation of Gödel's First Incompleteness Theorem from the Liar Paradox', *Proceedings of the 23rd Annual International Joint Conference on Artificial Intelligence (IJCAI-13)* .

Markman, A. & Gentner, D. (1993), 'Splitting the Differences: A Structural Alignment View of Similarity', *Journal of Memory and Language* **32**, 517–535.

Russell, S. & Norvig, P. (2009), *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ. Third edition.