By *Dis*analogy, Cyberwarfare is Utterly New

Selmer Bringsjord
Rensselaer AI & Reasoning (RAIR) Lab
Department of Cognitive Science
Department of Computer Science
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA
selmer@rpi.edu

John Licato
Rensselaer AI & Reasoning (RAIR) Lab
Department of Cognitive Science
Department of Computer Science
Rensselaer Polytechnic Institute (RPI)
Troy, NY 12180 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, AI, future

Contents

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 axiom systems for such things as arithmetic, from which at least the vast majority of arithmetic can be deductively derived.[1] 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: For 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, it is difficult to see how such a claim can be made without crucially relying on 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.[2] 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.[3] If Just War Theory (JWT) provides verdicts with respect to familiar

[1] For instance, Peano Arithmetic is a well-known and well-understood example of such a system. For an introduction, the reader can consult (Ebbinghaus, Flum & Thomas 1994).

[2] Consider e.g. the highly influential, commendable, and clearly well-intentioned *Tallinn Manual* (2013) (full title from CUP: *Tallinn Manual on the International Law Applicable to Cyber Warfare*), which explicitly seeks to transfer, by analogy and not by deduction or induction, both (jurisprudential versions of) *jus ad bellum* and *jus in bello* from the domain of conventional warfare to cyber-warfare. It is noteworthy, by the way, that the chief reason given by the authors of *TM* for constructing this volume was their perception of the — as they put it — "normative ambiguity" of cyberwarfare (Schmitt 2013 p. 17).

[3] Bringsjord has in fact attempted to show that when spear phishing is formalized in certain ways it turns out to be a special case of the age-old "haversack ruse." For the informal

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?

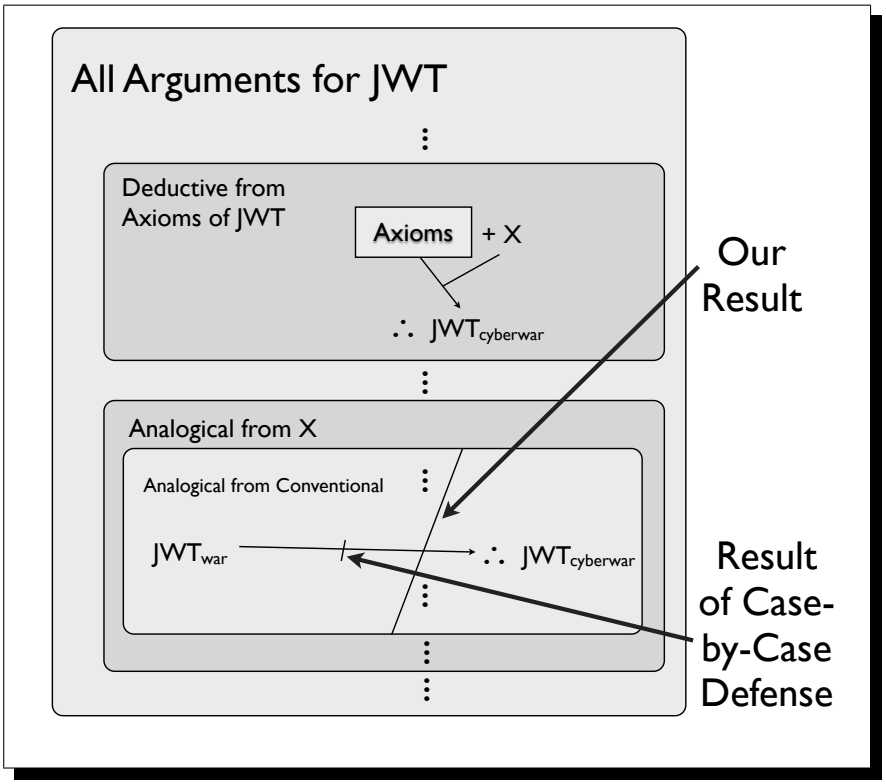

Figure 1: The argument we outline in the present paper aims to show not that some *particular* argument by analogy fails to hold between these two domains, but rather that these domains are dissimilar to the point that they cannot possibly produce *any* good analogy. This figure diagrams the space of analogical arguments to show that we intend not to refute one specific analogical argument between JWT_{war} and $\text{JWT}_{cyberwar}$, but rather *all* such arguments between those domains.

We believe that cyberwarfare (along with some forms of "mild" cybercon-

basis for Bringsjord's formal claim, see (Macintyre 2011).

flict) is not only fundamentally new, but, upon closer inspection, *dangerously* new. In order to defend our position, we could merely do 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.[4] 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 since this approach would need to disarm each individual case as it comes along, it's surely a rather inefficient approach. Moreoever, the inefficiency of this piecemeal approach, by our lights, marks it as unwise, because while cyberwarfare is still only in its infancy, its rapid growth (in part for reasons we canvass below) is inevitable. It would presumably be imprudent to allow plodding ethical analysis to be outsripped by the activity that that analysis is intended to cover.

Instead, we shift from the defensive to the offensive mode. We analyze traditional warfare, espionage, and cyberwarfare, producing not an axiomatic system for each that is poised to serve as a source of deduction, but producing instead a representative quartet of necessary conditions (for each concept) sufficient to undergird rigorous disanalogical reasoning. Next, via (deductive) meta-reasoning (over an instantiation of an underlying theory of argument) that shows *dis*analogy, we show that cyberwarfare — since its environment is not kinetic and is suffused with autonomous AIs — 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.[5] 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[6] had a stunningly long run, but today's world, based as it is on digital information and increasingly intelligent and autonomous 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.[7]

[4] Such an approach could e.g. be undertaken in connection with line-by-line analysis and assessment of the aforementioned *Tallinn Manual*.

[5] 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.

[6] And their likeminded predecessors, of course. E.g., clearly 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 cyberwarfare. Accordingly, we don't spend time rehearsing the roots of JWT, which are likely to be familiar to all of our readers.

[7] 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.

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 realm 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, the caption for which encapsulates the heart of that argument.) We then (§6) present and rebut four objections. Some concluding remarks close the paper.

2. Prior Work on Analogical Reasoning

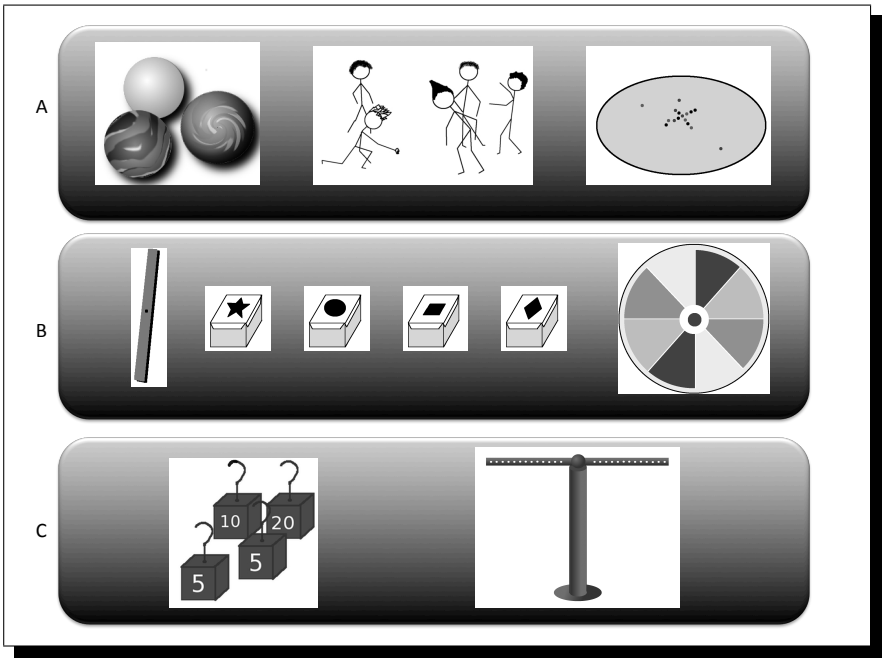

Figure 2: Objects and Concepts for a "Piaget-MacGyver Room" Experiment. See Bringsjord and Licato (2012) for more details.

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 et al. 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 formal 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.

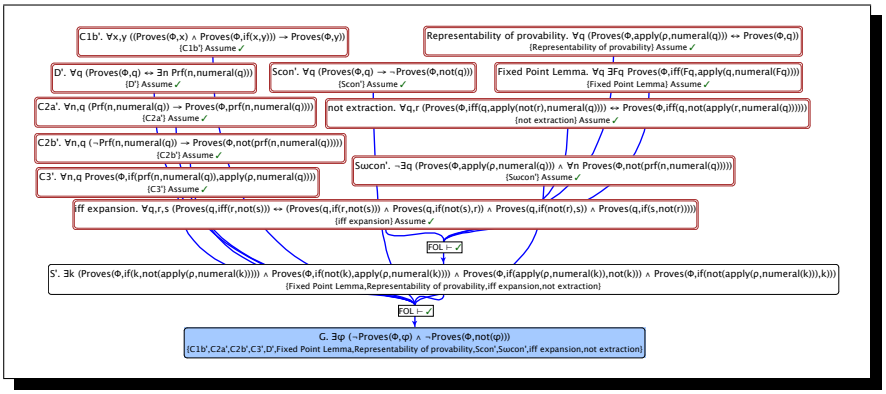

Figure 3: Full Deductive "Short-Distance" Proof of **G1** in Slate, Automatically Generated. See Licato et al. (2013) for more details.

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.

ADR is common in human reasoning, appearing, as we have pointed out, in the reasoning of children (Licato, Bringsjord & Hummel 2012, Bringsjord & Licato 2012), professional logicians (Licato, Govindarajulu, Bringsjord, Pomeranz & Gittelson 2013), and everywhere in between (Hummel, Licato & Bringsjord 2014). 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 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.[8] 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

[8] 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).

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 to the positive analogy $\mathcal{P}$, the schema also contains a negative analogy $\mathcal{A}, \mathcal{B}$. 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 (*proposals*) that serves as a placeholder for any additional conditions that are potentially relevant; that is, they either strengthen or weaken the hypothetical analogical inference χ. *proposals* contains any conditions that are, to use the language of Keynes (1921) and Bartha (2013), not in the positive, negative, neutral, or hypothetical analogy, but may yet have some effect on the argumentative force of the hypothetical analogy. We acknowledge here that there may be various alternate proposals by philosophers for the categorization of statements of relevance, including those which may prefer *proposals* to be subsumed under either the positive analogy $\mathcal{P}$ or the negative analogy $\mathcal{A}, \mathcal{B}$, but we leave it as a separate line in our schema $\mathscr{A}$ just in case such proposals are not universally accepted. In any case, 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, which are thus left to sedulous readers to investigate.[9]

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

[9]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 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^*

Table 2: Instantiation $\mathscr{A}^{w\rightarrow c}$ 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*
$\mathbf{JWT}_{\text{war}}$		
		$\mathbf{JWT}_{\text{cyberwar}}$

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, for 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 $\mathbf{JWT}_{\text{cyberwar}}$ 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. 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 $\mathscr{A}^{w \rightarrow c}$ (which 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 a conventional, purely kinetic gun in a firefight may involve a detailed chain of physical processes (the gunpowder firing, the bullet spinning) found in the SOURCE, but does not necessarily contain an analog 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 infers that canines aren't morally obligated to perform any actions. Since, as our thinker explains, parallel reasoning holds for all those falling under *Canis lupus*, and then indeed for mere animal after mere animal regardless of the species in question, $\bar{E}$ 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, where an attribute or property of a domain is essential if it is "causally closely related to the positive analogy" (Hesse 1966). However, as Bartha (2013) points out, Hesse's formulation of this so-called No-essential-difference condition may seem too restrictive, as it fails to acknowledge that even when the same objects are used in two separate analogical arguments, the attributes of those objects which are considered essential may differ. In other words, the essential attributes differ depending on the hypothetical analogy χ!

Our resolution is twofold: First, we stress that the interpretation of the schema in Table 1 used in this paper is based on Bringsjord's *Platonic* reading of Hesse's No-essential-difference condition. On this interpretation, domains and their objects therein have essential attributes that remain essential irrespective of any particular analogical argument in which they may appear.[10] Second, we acknowledge that this Platonic reading of Schema $\mathscr{A}$ makes the present paper's central argument *conditional*: The argument holds only if the schema in question is a normative structure that isn't compromised by any context- or person-relative factors. The required condition here does seem extremely plausible. While $\mathscr{A}$ may not be as uncontroversial as those normative schemas found in the classical-and-Platonic deductive sphere (e.g., *reductio ad absurdum*, which of course is questioned in constructive mathematics), certainly $\mathscr{A}$ should be accorded the same firm footing as schemas in, say, probability logic. The dominant view, for instance, is certainly that schemata such as the following Kolmogorovian one, covered in (Adams 1998) (and indeed in any orthodox presentation of probability logic), holds across all particular cases.

$$\text{If } \{\phi\} \vdash \psi, \text{ then } p(\phi) \leq p(\psi).$$

Here, 'p' denotes the probability function.[11]

[10] E.g., to be a person is to have certain attributes, and anything that is person *must* have these attributes; this is a fact wholly independent of any analogical argumentation one might make. [A definition of personhood is offered in (Bringsjord 1997)].

[11] It may alternately be preferable, as this paper's second author is inclined to do, to define

Now, with this definition of essential attributes in place, let us restate principle $\mathcal{E}$ as follows:

$\mathcal{E}$ If any of the essential attributes of the SOURCE are in the negative part of the relevant instance of the schematic argument $\mathscr{A}$, that instance fails.

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 speaking. Put another way, mere animals aren't *persons*, yet it is the qualities that constitute personhood which hold sway in moral matters. In our disproof, likewise, we will show that essential attributes of the SOURCE are indeed absent in the TARGET. Again, the mismatch between the essential attributes of the SOURCE and TARGET will undermine the validity of any analogical arguments that rely on those essential properties.[12] 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*.[13]

- *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 at least generally accessible to human cognition.

the essential attributes as those things which must hold in order for the formulae in the positive analogy $\mathcal{P}$ to hold. This option would be affirmed either because these attributtes logically follow, or are in some other way causally related (though they may not be explicitly specified in the positive analogy). Stated this way, the category of *essential attributes* starts to sound like the category of *prior associations* used in Bartha's (2010) Articulation Model, and affirming this line would further weaken the central claim of the present paper to one which only rules out arguments of analogy between $\mathbf{JWT}_{war}$ and $\mathbf{JWT}_{cyberwar}$ *if those arguments have the essential features (listed in Section 5.1) in their positive analogy.*

[12] This is to put the situation circumspectly. If we were to craft a more formal analysis, it would presumably be possible to *prove* that the deontic operators that would be present in the formulae that express $\mathbf{JWT}_{\text{war}}$ quite literally cannot apply in the AI-infused world brought on by what we below call The Minimaxularity. Such operators, in a quantified logic, which is what we have invoked above, require that the "slots" over which they range must necessarily be of the correct type.

[13] 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.

- *Proportionality.* This familiar set of principles dictates that in just war no attacks can be disproportional to the ends sought. This in turn entails that warfighters have an understanding of the effects of the actions they can perform, since without such 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 to distinguish between innocent non-combatants versus combatants, and that warfighters refrain from intentionally harming non-combatants in the course of seeking military victory.[14] 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, partial 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 — as the story goes — move beyond that level, and then exploit their superhuman powers to built smarter and smarter and ... smarter machines, leaving us in the dismal dust. The main argument for the proposition that The Singularity 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);

[14] This is to express the situation with an artificial narrowness. As e.g. von Clausewitz (1976) explains, war is frequently conducted to secure not victory, but leverage, including leverage in the political sphere.

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" (Bringsjord & Bringsjord 2015).[15] 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 or human-level creativity[16]), yet "*max*imal" with respect to certain logico-mathematical constraints.[17] These constraints, put rather impressionistically, which is a mode 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.[18]

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.[19] 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, 50 years from now with respect to intelligent agents. By 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 any number of nasty things that are today done manually by clever humans — 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

[15] Which is in turn forthcoming in the book *The Technological Singularity: A Pragmatic Perspective* (Miller, Callaghan, Yampolskiy, Armstrong & Lahee 2015). For the position that The Singularity is a fideistic pipe-dream, see (Bringsjord, Bringsjord & Bello 2013).

[16] Indeed, S. Bringsjord and A. Bringsjord hold that computing machines *cannot possibly* have phenomenal consciousnessness (Bringsjord 1992) or be genuinely creative (Bringsjord, Ferrucci & Bello 2001).

[17] This distills to the vision explicated in (Bringsjord 1992).

[18] This can be put another way, with help from the dominant, encyclopedic handbook of modern AI: The MiniMaxularity is the future point at which the techniques in *Artificial Intelligence: A Modern Approach* (2009) are not only further refined (without the introduction of fundamentally new formal paradigms), but are run on hardware that is many orders of magnitude faster than what is available today, *and* are fully integrated with each other and placed interoperably within particular artificial agents.

[19] For an overview of Watson, see (Ferrucci et al. 2010). For an account of the logico-mathematical nature of Watson, in the context of Leibniz's vision for rational cognition, see (Bringsjord & Govindarajulu forthcoming).

smart and autonomous 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; see §6.4.) 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 flat-out impossible to access ordinary kinetic weapons without first specifically engaging AIs that are inextricably bound up in these cyber layers. The layers of software that will enshroud all things analog will in our view 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 is easy-as-pie to obtain, and we give here only the informal proof-sketch:

> **Proof-sketch**: We first simply note again 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. This will for instance mean that there will be no such thing as today's "pulling a trigger." Hence, 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 principle $\mathcal{E}$, 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.[20] 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.

[20] A profitable paper to study in this regard is (Taddeo 2012). This paper, by the way, which came to our attention after the first and second drafts of the present paper were composed, anticipates the problem of discrimination for cyberwarfare, by seeing that problem as one plaguing *information* warfare.

6. Objections

In this section we consider and rebut two objections to our case. The first is philosophical and prudential in nature; the second is technical.

6.1. Objection 1: Bypassing Autonomy?

"I understand your assumption that the MiniMaxularity will soon come to pass, from which it follows that an era is fast approaching in which artificial agents will have not only considerable intelligence, but autonomy as well. This state-of-affairs I accept; everyone should. What I *don't* accept is something more specific that your disproof needs: namely, that the artificial agents in the 'MiniMax' era will be deployed in warfare, specifically in cyberwarfare. I reject this premise, because clearly the deployment of such agents could 'boomerang.' Furthermore, I note that you nowhere establish the premise in question. In fact, you don't say *anything* in support of this proposition."

Reply: As I (Selmer) type this sentence, I'm listening to a presentation in Washington DC on a research project sponsored by the Office of Naval Research devoted to the engineering of autonomous artificial agents able to assist human warfighters. Now, it's true that the agents in this case are to have no kinetic power at all, but my first point in rebuttal is simply this undeniable two-fold one: Autonomous artificial agents *are* being built by a NATO country (viz. the US); some of these agents are designed for defensive purposes in the cyberwarfare realm.

Now for the second point, which, given implicit premises that I won't bother to unearth here, follows from the first: Countries/groups outside NATO are likewise seeking to build autonomous artificial agents for warfare, including specifically *cyber*warfare.

And now for the third point: Sooner or later, unfortunately probably sooner, an enemy of NATO is going to engineer and deploy autonomous artificial agents in order perpetrate significant *harm*. Does it not follow commonsensically from this third proposition that the US, or NATO overall, or perhaps first another particular country within NATO, will then have no choice but to fight fire with fire, and deploy autonomous artificial agents, operating at a time scale inaccessible to the relatively slow-moving decision-making power of *Homo sapiens sapiens*? If we only had propeller planes, and the enemy jets, you can bet your bottom dollar that we'd seek to invent and deploy the latter. The same logic holds with respect to autonomous machines. Even if we decide now to refrain from building autonomous machines for offensive purposes in warfare, an enemy, sooner or later, will attack us with machines able to achieve leverage greater than what non-autonomous machinery can deliver. In short, put starkly, it seems obvious that the age of MiniMaxularity will be one in which, just as we have presupposed in our disproof, autonomous artificial agents will be unleashed into a world where

nearly everything is digitally connected and shrouded.[21]

6.2. Objection 2: Key Principle Problematic?

"Your principle $\mathcal{E}$, which plays a crucial role in your purported disproof, is false! In fact, Bartha (2013) himself gives all we need to see that this principle fails, because he asks us to consider an example (#7 in his list of examples) of an apparent analogy between a 2-D rectangle and a 3-D box; he writes: 'Suppose that you have established that of all rectangles with a fixed perimeter, the square has maximum area. By analogy, you conjecture that of all boxes with a fixed surface area, the cube has maximum volume.'

Reply: We commend our opponent for scholarly alertness, but when one looks formally at the details, the objection evaporates. To see this, note that, expressed in our machinery, introduced and explained above, χ, with predicates that we charitably introduce in order to rigorize our opponent's point, is instantiated to:

$$\chi_r \quad \forall x, y, z[(Rect(x) \wedge Per(x, y) \wedge MaxArea(x, z) \wedge Area(x, z)) \rightarrow Square(x)]$$

And what is χ_r^*? This parallel formula:

$$\chi_r^* \quad \forall x, y, z[(Cuboid(x) \wedge SurfArea(x, y) \wedge MaxVolume(x, z) \wedge Area(x, z)) \rightarrow Cube(x)]$$

But now we see that Objection 2 does indeed evaporate. Why? Clearly, in *both* the 2-D and 3-D environments, involving in the former case rectangles, and in the latter case (of polyhedra) cuboids, the general property that is firmly operative in both environments is maximum space: space enclosed by polygons in 2-D environments, and space enclosed by polyhedra in 3-D environments.

Notice that what we have here is radically different than what we observed in the case of mere animals vs. humans considered in section 5: Whereas the concept of 2-dimensionality is not the underlying reason for the relevant instantiation of χ, and indeed this property of 2-dimensionality doesn't appear in χ_r, the situation is the opposite in the animal-human case, where the sub-human cognitive capacity constitutive of mere animals is precisely what precludes mere animals having the status of a true moral agents, and a biological correlate to maximum space is nowhere to be found.

[21] Our rebuttal to Objection 1 should not be read as implying that our position is that ethical countries will choose to engage in (future) cyberwar. We are simply explaining that the introduction of autonomous machines into cyberwarfare (and, for that matter, kinetic warfare), on the part of NATO, is in the process of happening, and that this trend will inexorably and inevitably continue. Our explanation serves our dialectical purposes at hand, and any consideration of the ethical status of those countries that introduce autonomous machines is a separate, orthogonal discussion relegated to another time and place.

6.3. Objection 3: Entirely Suppositional?

"Your argument is suppositional; it supposes a future world wherein the MiniMaxulaity has been reached. As such, any conclusion drawn is conditional in nature (whether material conditional or other). Thus, you tell us nothing about the nature of the world/cyberwarfare as we know it today, but only about the nature of the world/cyberwarfare as it might be if/when your highly speculative predictions about the future are bore out. With this in mind, your paper might be more accurately titled: 'By Disanalogy, Future Cyberwarfare will be Utterly New.'

Reply: Our argument is indeed suppositional; we have made no secret of this whatsoever. The mere fact that some argument is suppositional in nature doesn't constitute a flaw in that argument. More precisely, the mere fact that some argument's conclusion rests on a supposition[22] in that argument does nothing to impugn the argument. In fact, this holds not just for arguments; it holds for *proofs*. Every number-theoretic theorem that has been proved rests on an infinite number of unstated assumptions — that the Peano Axioms hold, that axiomatic set theory is consistent, and so on *ad infinitum*.

Therefore, Objection 3 only has traction insofar as the assumptions our critic has in mind are false, or at least implausible. But it should be obvious to everyone that the MiniMaxularity (*not*, as we took pains to state, The Singularity!) is fast coming — for the simple reason that it can literally be *seen*. Now, yes, there are existential threats to planet Earth; this we all know. The sun might not rise tomorrow where our critic resides, because — to mention again a possibility noted in §5.2 — Earth is (in his/her) overnight vaporized by an asteroid. In this case, even the foreseeable future to which we have pointed will not arrive. But barring the arrival of such an asteroid (or a tidal wave that sweeps away all of civilization, or the arrival of advanced aliens who throw us back to the Stone Age, or a "superbug" pandemic, etc.), we can see what is coming. Since the advent of the logico-mathematics of computation (in the form of the λ-calculus, or Post machines, or your favorite formalism), combined with physicalization and harnessing of these abstractions, AI has been on a steady, ascending journey that gives birth to increasingly intelligent artificial agents (the lastest famous case being Watson, cited above). There are parts of personhood that no mere computing machine can ever duplicate (as e.g. explained and shown in Bringsjord 1992), but the MiniMaxularity is based simply on a sober observation of what is happening before our very eyes. When one sits in a self-driving car from Google, a rational human should of course not issue the prediction that taxi drivers will be disemployed in 2016; but sitting in such a car should allow any sensible person to see that, assuming, say, a half-century of further AI/robotics progress, human taxi drivers

[22] Or, better — since in so-called *natural* deduction suppositions are *discharged*; see e.g. (Barwise & Etchemendy 1999), or any standard and elementary coverage of such deduction —, an *assumption*.

won't be needed in 2066. Instead, smart software will drive us to and fro, and likewise transport any and all inanimate things we wish moved considerable distances, at high rates of speed.

6.4. Objection 4: Enshrouding Hogwash?

"A crucial part of your case is the premise that AI agents will, as you put it, 'enshroud' standard kinetic causation within a digital world. With all due respect, this is simply hogwash.

Standard kinetic causation will remain. It is true that (future) cyberwar and enshrouding autonomy technology will increase the causal *distance* between humans and events in the world, and there will be a diminished ability to predict the consequence of events. This will affect the determination of what is permissible or not under JWT, but does not affect the applicability of JWT. The claim that essential attributes "fail to hold" is only sensible as 'fail to be true;' the attributes still apply. Thus, we may be left with a result that seemingly analogous physical/cyber acts are not equally just under JWT because the analogy between the acts is flawed (e.g., they do not have correspondingly equal proportionality, discrimination, etc). However, JWT is not threatened."

Reply: Actually, we don't claim that the enshrouding in question is by AI agents, which is what the skeptic here asserts. Rather, we say only that *digital software* will enshroud physical objects. Of course, some, perhaps most (indeed, perhaps, for all we know, all) of this software will be intelligent; who knows. Regardless, we in no way employ any such premise as that AIs specifically do the enshrouding.[23]

Despite the skeptic's stern incredulity, the fact is that the enshrouding we foresee, according even to non-academicians deeply acquainted with and invested in the commerce of real-world software and hardware, is *already* underway, readily observable, quite disturbing, and apparently unstoppable. A well-known example is Marc Andreessen, who has famously explained that software is "eating the world" (Andreessen 2011). While we of course in no way can be fairly interpreted as asserting or presupposing or implying that physical things and physical causation will vanish, the fact is that some highly respected and relevant thinkers see coming a revolutionary and radical degree of enshrouding — to the point, for example, where Floridi (n.d.) perceives an information revolution, now unfolding, in which our ordinary world is, to use his instructive term, "de-physicalized." Taddeo (2012) affirms Floridi's diagnosis and vision, and a paper quite relevant to the present one, provides a discussion of "information" warfare in the context of a Floridian future. These thinkers, and no doubt others, would agree that a time is fast approaching when the simplest purely physical causal chains of today will be mediated by

[23] We do opine that the shrouding software will corrrespond to AIs; see §5.2.

software.

Finally, our critic might be well-served by studying the theory of war, as expounded by relevant seminal thinkers. For these thinkers work with a conceptual model based thoroughly on standard physical causation and ordinary material things. For instance, von Clausewitz (1976), in his landmark *On War*, says explicitly that war is fundamentally nothing more than a "dual" on an "extensive scale."[24]

7. 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.

[24] As to the critic's distinction here between 'fail to hold' vs. 'fail to be true,' we dealt with this issue in note # 12.

Many thanks to NATO CCDCOE for the opportunity to share our ideas in Rome at the November 2013 Workshop on Ethics and Policies for Cyber Warfare. We thank William Boothby for spirited objections, expressed in person in the Eternal City, to our disproof and our rather dystopian view of the future. Special thanks are due to both Mariarosaria Taddeo and Luciano Floridi for numerous insights, the production of which, for them, is par for a rather elevated course. Two anonymous referees provided incisive objections, comments, and corrections, and we express deep gratitude to them as well. Finally, the very concept of the MiniMaxularity, toward which our race is racing, is originally due to another, younger Bringsjord.

References

Adams, E. (1998), *A Primer of Probability Logic*, CSLI, Stanford, CA.

Andreessen, M. (2011), 'Why Software is Eating the World', *Wall Street Journal* . The piece appeared in the *Life & Culture* section.

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.

Barwise, J. & Etchemendy, J. (1999), *Language, Proof, and Logic*, Seven Bridges, New York, NY.

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. (2015), The Singularity Business, *in* J. Miller, V. Callaghan, R. Yampolskiy, S. Armstrong & A. Lahee, eds, 'The Technological Singularity: A Pragmatic Perspective', Springer, New York, NY.

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., Ferrucci, D. & Bello, P. (2001), 'Creativity, the Turing Test, and the (Better) Lovelace Test', *Minds and Machines* **11**, 3–27.

Bringsjord, S. & Govindarajulu, N. S. (forthcoming), Leibnizs Art of Infallibility, Watson, and the Philosophy, Theory, & Future of AI, *in* V. Müller, ed., 'Synthese Library', Springer.
URL: *http://kryten.mm.rpi.edu/SB_NSG_Watson_Leibniz_PT-AI_061414.pdf*

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.

Ebbinghaus, H. D., Flum, J. & Thomas, W. (1994), *Mathematical Logic (second edition)*, Springer-Verlag, New York, NY.

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*

Floridi, L. (n.d.), 'The Digital Revolution as the Fourth Revolution', *Digital Revolution* . This was a contribution to the online program ("Digital Revolution") from BBC.

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.

Hummel, J. E., Licato, J. & Bringsjord, S. (2014), 'Analogy, explanation, and proof', *Frontiers in Human Neuroscience* **8**(867).

Keynes, J. (1921), *A Treatise on Probability*, Macmillan, London.

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)* .

Macintyre, B. (2011), *Operation Mincemeat: How a Dead Man and a Bizarre Plan Fooled the Nazis and Assured an Allied Victory*, Broadway Books, New York, NY.

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

Miller, J., Callaghan, V., Yampolskiy, R., Armstrong, S. & Lahee, A., eds (2015), *The Technological Singularity: A Pragmatic Perspective*, Springer, New York, NY. This book is the second in a two-volume set on The Singularity from Springer. The first is *The Singularity Hypothesis*, edited by A. Eden, J. Moor, J. Søraker and E. Steinhart (2013).

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

Schmitt, M., ed. (2013), *Tallinn Manual on the International Law Applicable to Cyber Warfare*, Cambridge University Press, Cambridge, UK. This volume was first published in 2011. While M Schmitt is the General Editor, there were numerous contributors, falling under the phrase 'International Group of Experts at the Invitation of the NATO Cooperative Cyber Defence Centre of Excellence'.

Taddeo, M. (2012), 'Information Warfare: A Philosophical Perspective', *Philosophy of Technology* **25**, 105–120.

von Clausewitz, C. (1976), *On War*, Princeton University Press, Princeton, NJ. This is a translation, by M. Howard and P. Paret, of the 1984 edition from Princeton University Press. *On War* was originally published in 1832.