فرضية العالم المعرض للخطر

The Vulnerable World Hypothesis

Nick Bostrom, trans. Dr Hussein Abdul Ghani Ebraheem

Original: [Global Policy, Vol. 10, No. 4 (2019): 455–476] [pdf]

بروفيسور/ نيك بوستروم (*) معهد مستقبل الإنسانية ، جامعة أكسفورد

ترجمة دكتور/ حسين عبدالغني إبراهيم ، باحث حر ، مصر

نبذة مختصرة **Abstract**

قد يغير التقدم العلمي والتكنولوجي من قدرات الناس أو دوافعهم بطرق من شأنها زعزعة استقرار الحضارة. على سبيل المثال ، التقدم في أدوات الاختراق البيولوجي (**) DIY قد يجعل من السهل على أي شخص لديه تدريب أساسي في علم الأحياء قتل الملايين ؛ فيمكن للتكنولوجيات العسكرية الجديدة أن تطلق سباقات تسلح تكون الميزة الحاسمة فيها لمن يضرب أولاً ؛ أو قد يتم اختراع بعض العمليات المفيدة اقتصاديًا والتي تنتج عوامل خارجية سلبية كارثية من الصعب تنظيمها. تقدم هذه الورقة مفهوم العالم المعرض للخطر valnerable world : تقريبًا باعتباره ، عالمًا يوجد فيه مستوى معين من التطور التكنولوجي حيث يتم تدمير الحضارة بشكل افتراضي تقريبًا ، أي ما لم تكن قد خرجت من "حالة التقصير شبه الفوضوية". يجري تحليل عددا من نقاط الخطورة التاريخية المحتملة والمضاربة المستقبلية وترتيبها في ملف التصنيف. إن القدرة العامة على تحقيق الاستقرار في عالم معرض للخطر تتطلب قدرات متضخمة إلى حد كبير للشرطة الوقائية والحوكمة العالمية. وهكذا تقدم فرضية العالم المعرض للخطر منظورًا جديدًا يمكن من خلاله تقييم توازن المخاطر والفوائد للتطورات باتجاه المراقبة الشاملة أو النظام العالمي أحادي القطب.

التأثيرات السياسية **Policy Implications**

– لا ينبغي لسياسة التكنولوجيا أن تفترض ، بصورة لا شك فيها ، أن كل التقدم التكنولوجي مفيد، أو أن كل فتح علمي هو الأفضل دائمًا وبصورة كاملة ، أو أن العالم لديه القدرة على إدارة أي جانب سلبي محتمل للتكنولوجيا بعد اختراعها.

– بعض المجالات ، مثل البيولوجيا الصناعية ، يمكن أن تنتج اكتشافًا يؤدي فجأة إلى تبرير الدمار الشامل ديمقراطيًا ، على سبيل المثال من خلال تمكين الأفراد من قتل مئات الملايين من الأشخاص باستخدام مواد متاحة بسهولة. ومن أجل أن تتمتع الحضارة بقدرة عامة على التعامل

مع اختراعات "الكرة السوداء black ball" من هذا النوع ، فإنها سوف تحتاج إلى نظام مراقبة عالمية شاملة في الوقت الحالي. في بعض السيناريوهات ، يجب أن يكون مثل هذا النظام في مكانه قبل اختراع التكنولوجيا.

– يمكن توفير حماية جزئية ضد مجموعة محدودة من الكرات السوداء المحتملة من خلال تدخلات أكثر استهدافًا. فمثلا، يمكن التخفيف من المخاطر البيولوجية عن طريق رصد العاملين والتحقق من خلفياتهم في بعض أنواع معامل البيولوجيا ، وعن طريق تثبيط الاختراق البيولوجي (على سبيل المثال من خلال متطلبات الترخيص) ، وإعادة هيكلة قطاع التكنولوجيا الحيوية للحد من الوصول إلى بعض الأجهزة والمعلومات المتطورة. وبدلاً من السماح لأي شخص بشراء آلة تخليق الحمض النووي الخاصة به ، يمكن توفير تخليق الحمض النووي كخدمة من قبل عدد صغير من مقدمي الخدمة الخاضعين للمراقبة عن كثب.

– نوع آخر من الكرة السوداء أكثر دقة هو النوع الذي يعزز حوافز الاستخدام الضار – على سبيل المثال. تقنية عسكرية تجعل الحروب أكثر تدميراً مع إعطاء ميزة أكبر للجانب الذي يضرب أولاً. فيجب أن تُستخدم أوقات السلام النسبي لبناء آليات أقوى لحل النزاعات الدولية ، مثل السنجاب الذي يستخدم أوقات الوفرة لتخزين الجوز لفصل الشتاء.

هل هناك كرة سوداء في جرة/جعبة الاختراعات المحتملة؟

Is there a black ball in the urn of possible Inventions?

طريقة واحدة للنظر إلى الإبداع البشري هي النظر إليه باعتباره عملية سحب الكرات من جرة عملاقة. (١) الكرات تمثل الأفكار ، والاكتشافات ، والاختراعات التكنولوجية الممكنة. لقد استخرجنا عددًا كبيرًا من الكرات على مدار التاريخ – في الغالب كرات بيضاء (مفيدة) ولكن توجد أيضًا ظلال مختلفة من الرمادي (معتدلة الضرر وأخرى تمزج بين النعم والنقم). لقد كان التأثير التراكمي على الحالة البشرية إيجابيًا للغاية حتى الآن ، وربما يكون أفضل بكثير في المستقبل (Bostrom, 2008) . فقد نما عدد سكان العالم من حيث الحجم بنحو ثلاثة أضعاف على مدى العشرة آلاف سنة الماضية ، وفي القرنين الماضيين ، ارتفع دخل الفرد ومستويات المعيشة ومتوسط العمر المتوقع أيضًا. (٢)

ما لم نستخرجه ، حتى الآن ، هو الكرة السوداء: التقنية التي تدمر بشكل ثابت أو افتراضي الحضارة التي تخترعها. السبب ليس أننا كنا حذرين أو حكماء بشكل خاص في سياستنا التكنولوجية. لقد حالفنا الحظ فقط.

لا يبدو أن ثمة حضارة إنسانية جرى تدميرها – بما يعوق تحولها – بواسطة اختراعاتها.(٣) لدينا بالفعل أمثلة على حضارات تم تدميرها من خلال اختراعات مصنوعة في أماكن أخرى. على سبيل المثال ، يمكن النظر إلى الاختراعات الأوروبية التي مكنت السفر عبر المحيط والاعتراض بالقوة على أنها حدث كرة سوداء للسكان الأصليين في الأمريكتين وأستراليا وتسمانيا Tasmania وبعض الأماكن الأخرى. كما أن انقراض مجموعات البشر القديمة ، مثل إنسان الـ نياندرتال والـ دينيسوفان ، ربما تم تسهيله من خلال التفوق التكنولوجي للإنسان العاقل Homo Sapiens . ولكن حتى الآن ، على ما يبدو ، لم نشهد أي اختراع مدمر ذاتيًا بما يكفي ليتم اعتباره كرة سوداء للبشرية. (٤)

ماذا لو كانت هناك كرة سوداء في الجرة؟ إذا استمر البحث العلمي والتكنولوجي ، فسوف نصل إليها في النهاية ونخرجها. إن حضارتنا لديها قدرة كبيرة على التقاط الكرات ، ولكن لا توجد قدرة على إعادتها إلى الجرة مرة ثانية. إننا يمكننا أن نخترع ولكن لا يمكننا أن نقوم بالفعل الذي يُضاد عملية الاختراع un–invent . فاستراتيجيتنا هي الأمل في عدم وجود كرة سوداء.

هذه الورقة تطور بعض المفاهيم التي يمكن أن تساعدنا في التفكير في إمكانية وجود كرة تكنولوجية سوداء ، والأشكال المختلفة التي يمكن أن تتخذها هذه الظاهرة. نناقش أيضًا بعض الآثار المترتبة على السياسة من منظور عالمي ، لا سيما فيما يتعلق بالكيفية التي يجب أن ينظر بها المرء إلى التطورات في مجال المراقبة الجماعية والتحركات نحو حوكمة عالمية أكثر فعالية أو نظام عالمي يفوق أحادية القطب. هذه الآثار لا تحسم بأي حال من الأحوال الأسئلة حول الرغبة في إجراء تغييرات في تلك المتغيرات الاستراتيجية الكلية – لأن هناك بالفعل عوامل أخرى وثيقة الصلة ، لم يتم تناولها هنا ، والتي يجب إضافتها إلى التوازن. ومع ذلك فهي تشكل مجموعة مهمة من الاعتبارات التي لا تحظى بالتقدير الكافي والتي ينبغي أن تؤخذ في الاعتبار في المناقشات المستقبلية حول هذه القضايا.

قبل الوصول إلى الأجزاء الأكثر مفاهيمية في الورقة ، سيكون من المفيد رسم صورة أكثر واقعية لما يمكن أن تكون عليه الكرة السوداء التكنولوجية. النوع الأكثر وضوحًا هو التكنولوجيا التي من شأنها أن تجعل من السهل جدًا إطلاق العنان لقوة تدميرية هائلة. من الواضح أن التفجيرات النووية هي القوة الأكثر تدميراً التي نتقنها. لذا دعونا نفكر فيما كان سيحدث لو كان من السهل جدًا إطلاق هذه القوة.

تجربة فكرية: قنابل نووية سهلة **A thought experiment: easy nukes**

في صباح يوم ١٢ سبتمبر ١٩٣٣ ، كان ليو تسيلارد Leo Szilard يقرأ الجريدة عندما وجد تقريرًا عن خطاب ألقاه مؤخرًا لورد رذرفورد Lord Rutherford المميز ، الذي غالبا ما يُعتبر الآن والد الفيزياء النووية (Rhodes, 1986) . رفض رذرفورد ، في خطابه ، فكرة استخلاص الطاقة المفيدة من التفاعلات النووية ووصفها بأنها "لغو". هذا الادعاء أزعج تسيلارد لدرجة أنه خرج في نزهة على الأقدام. أثناء المشي ، خطرت له فكرة سلسلة من التفاعلات النووية – وهي أساس كل من المفاعلات النووية والقنابل النووية. أظهرت الأبحاث اللاحقة أن صنع سلاح نووي يتطلب عدة كيلوجرامات من البلوتونيوم plutonium أو اليورانيوم uranium عالي التخصيب ، وكلاهما صعب الإنتاج ومكلف للغاية. ومع ذلك ، لنفترض أن الأمر قد تبين عكس ذلك: وأن هناك طريقة سهلة حقًا لإطلاق طاقة الذرة – لنقُل ، عن طريق إرسال تيار كهربائي عبر جسم معدني موضوع بين لوحين من الزجاج.

لذلك دعونا نفكر في التاريخ المضاد الذي يخترع فيه تسيلارد الانشطار النووي ويدرك أن القنبلة النووية يمكن صنعها بواسطة قطعة من الزجاج ، وجسم معدني ، وبطارية مضبوطة ذات تكوين معين. ماذا يحدث بعد ذلك؟ يصبح تسيلارد قلقًا للغاية. ويرى أن اكتشافه يجب أن يظل سراً بأي ثمن. ولكن كيف؟ لا بد أن تصل رؤيته للآخرين. يمكنه التحدث إلى عدد قليل من أصدقائه الفيزيائيين ، والذين من المرجح أن يعثروا على الفكرة ، ويحاول إقناعهم بعدم نشر أي شيء عن سلسلة التفاعلات النووية أو عن أي من خطوات التفكير التي تؤدي إلى الاكتشاف الخطير. (هذا ما فعله تسيلارد في التاريخ الفعلي) .

هنا يواجه تسيلارد معضلة: إما أنه لا يشرح الاكتشاف الخطير ، لكنه لن يكون فعالاً في إقناع عدداً من زملائه بالتوقف عن النشر ؛ أو يخبرهم عن سبب قلقه ، لكنه ينشر بعد ذلك المعرفة الأكثر خطورة. في كلتا الحالتين يخوض معركة خاسرة. التقدم العام للمعرفة العلمية ، في نهاية المطاف ، سوف يجعل الرؤية الخطيرة أكثر سهولة. وبسرعة ، لن يتطلب اكتشاف كيفية بدء سلسلة تفاعل نووي باستخدام قطع من المعدن ، والزجاج ، والكهرباء ، لن يتطلب عبقرية ، ولكنه سيكون في متناول أي طالب لديه عقلية إبداعية في العلوم والتكنولوجيا والهندسة والرياضيات.

دعونا نلف الشريط قليلا. يبدو الوضع ميؤوسًا منه ، لكن تسيلارد لا يستسلم. ويقرر أن يضع ثقته في أحد الأصدقاء ، صديق هو أيضًا أشهر عالِم في العالَم – ألبرت أينشتاين. نجح في إقناع أينشتاين بالخطر. (مرة أخرى نتتبع التاريخ الفعلي) . الآن ، يحظى تسيلارد بدعم رجل يمكنه أن يرتب له جلسة استماع مع أي حكومة. فيما بعد ، يكتب الاثنان رسالة إلى الرئيس فرانكلين دي روزفلت. بعد بعض المشاحنات بين اللجان وكتابة التقارير ، أصبحت المستويات العليا في حكومة الولايات المتحدة مقتنعة في النهاية بما يكفي لتكون جاهزة لاتخاذ إجراءات جادة.

ما هو الإجراء الذي يمكن أن تتخذه الولايات المتحدة؟ دعونا أولاً نفكر في ما حدث بالفعل (Rhodes ، 1986) . ما فعلته حكومة الولايات المتحدة ، بعد استيعاب المعلومات التي قدمها آينشتاين وتسيلارد ، وبعد تلقي بعض التنبيهات الإضافية من البريطانيين الذين كانوا ينظرون أيضًا في الأمر ، هو إطلاق مشروع مانهاتن من أجل تسليح الانشطار النووي في أسرع وقت ممكن. وبمجرد أن أصبحت القنبلة جاهزة ، استخدمها سلاح الجو الأمريكي لتدمير المراكز السكانية اليابانية. برر الكثير من علماء مانهاتن مشاركتهم بالإشارة إلى الخطر المميت الذي قد ينشأ إذا حصلت ألمانيا النازية على القنبلة أولاً ؛ لكنهم استمروا في العمل على المشروع بعد هزيمة ألمانيا. (٥) دعا تسيلارد دون جدوى لتجريب "الأداة" على منطقة غير مأهولة بالسكان بدلاً من تجريبها على مدينة (Frank et al., 1945) . بعد انتهاء الحرب ، فضل الكثير من العلماء المراقبة الدولية للطاقة الذرية وأصبحوا ناشطين في حركة نزع السلاح النووي ؛ لكن وجهات نظرهم لم يكن لها وزن كبير ، حيث تم انتزاع السياسة النووية من أيديهم. بعد أربع

سنوات ، فجر الاتحاد السوفيتي قنبلته الذرية. ساعد الجواسيس من مشروع مانهاتن الجهود السوفيتية ، لكن حتى بدون التجسس كان سينجح في غضون عام أو عامين آخرين (Halloway, 1994) . تبعت ذلك الحرب الباردة ، التي شهدت في ذروتها ٧٠,٠٠٠ رأس نووي جاهزة لإطلاق العنان للدمار العالمي في أي لحظة ، بإصبع مرتجف يحوم فوق "الزر الأحمر" على أي من الجانبين (Norris and Kristensen, 2010) . (٦)

لحسن حظ الحضارة الإنسانية ، بعد تدمير هيروشيما وناجازاكي ، لم يتم تفجير أية قنبلة ذرية أخرى بدافع من الغضب. بعد ثلاثة وسبعين عامًا ، وبفضل المعاهدات الدولية وجهود منع انتشار الأسلحة النووية جزئيًا ، تمتلك تسع دول فقط أسلحة نووية. لا يُعتقد أن أي جهة من غير هذه الدول تمتلك أسلحة نووية على الإطلاق. (٧)

ولكن كيف كانت ستسير الأمور إذا كانت هناك طريقة سهلة لصنع الأسلحة النووية؟ هل كان من المحتمل أن يستطيع تسيلارد وأينشتاين إقناع حكومة الولايات المتحدة بحظر جميع الأبحاث في الفيزياء النووية (خارج المنشآت الحكومية شديدة الأمان)؟ مثل هذا الحظر على العلوم الأساسية كان سيتعرض لتحديات قانونية وسياسية هائلة – خاصة وأن سبب الحظر لا يمكن الكشف عنه علنًا بأي تفاصيل دون اختلاق معلومات غير مقبولة عن الخطر. (٨)

لنفترض ، مع ذلك ، أن الرئيس روزفلت كان بإمكانه ، بطريقة ما ، حشد دعم سياسي كافٍ للمضي قدمًا في فرض الحظر ، وأن المحكمة العليا الأمريكية يمكنها ، بطريقة ما ، أن تجد طريقة لاعتباره صالحًا دستوريًا. عندئذٍ نواجه مجموعة من الصعوبات العملية الهائلة. سيتعين إغلاق جميع أقسام الفيزياء بالجامعة ، وبدء الفحوصات الأمنية. عدد كبير من أعضاء هيئة التدريس والطلاب سيتم إجبارهم على الخروج. وكانت ستدور التكهنات الشديدة حول سبب كل هذه الإجراءات القاسية. كما كانت مجموعات من طلاب الدكتوراه وأعضاء هيئة التدريس في الفيزياء الذين تم منعهم من مجال أبحاثهم سوف يجلسون ويتأملون ما قد يكون الخطر السري. بعضهم كانوا سيكتشفون ذلك. ومن بين أولئك الذين اكتشفوه ، وسيشعر البعض بأنهم مضطرون لاستخدام المعرفة لإقناع زملائهم ؛ وهؤلاء الزملاء يريدون إخبار الآخرين ، لإثبات أنهم على دراية. بدلاً من ذلك ، قد يقرر الشخص الذي عارض الحظر من جانب واحد نشر السر ، ربما من أجل دعم وجهة نظره بأن الحظر غير فعال ، أو أن فوائد النشر تفوق المخاطر. (٩) (١٠)

الموظفون غير المبالين أو الساخطون في المختبرات الحكومية في النهاية كانوا سيتركون المعلومات تفلت من أيدينا ، وكان الجواسيس سيحملون السر إلى عواصم أجنبية. حتى لو ، بمعجزة ما ، لم يتم تسريب السر في الولايات المتحدة ، فإن العلماء في البلدان الأخرى كانوا سيكتشفونه بشكل مستقل ، وبالتالي مضاعفة المصادر التي يمكن أن ينتشر من خلالها. عاجلاً أم آجلاً – وربما عاجلاً – لن يكون السر سرًا بعد ذلك الوقت.

في العصر الحالي ، عندما يكون بإمكان المرء أن ينشر على الفور وبشخصية مجهولة على الإنترنت ، سيكون الحد من انتشار الأسرار العلمية أكثر صعوبة (Cf. Greenberg, 2012; Swire, 2015) .

قد يكون النهج البديل هو التخلص من جميع الزجاج أو المعدن أو مصادر التيار الكهربائي (ربما باستثناء عدد قليل من المستودعات العسكرية شديدة الحراسة) . نظرًا لوجود هذه المواد في كل مكان ، فإن مثل هذا التعهد سيكون شاقًا للغاية. لن يكون تأمين الدعم السياسي لمثل هذه الإجراءات أسهل من إغلاق تعليم الفيزياء. ومع ذلك ، بعد أن ارتفعت غيوم التفجيرات النووية فوق عدد قليل من المدن ، ربما يمكن حشد الإرادة السياسية للقيام بهذه المحاولة. ويعد استخدام المعادن مرادفًا تقريبًا للحضارة ، ولن يكون هدفًا واقعيًا للتخلص منها. كما يجب حظر إنتاج الزجاج ومصادرة الألواح الزجاجية الموجودة ؛ لكن قطع الزجاج ستبقى مبعثرة في جميع أنحاء الطبيعية لفترة طويلة. أيضا تجب مصادرة البطاريات والمغناطيسات ، على الرغم من أن بعض الأشخاص ربما خبأوا هذه المواد قبل أن يتم جمعها من قبل السلطات. وسيتم تدمير العديد من المدن من قبل العدميين ، والابتزازين ، والانتقاميين ، أو حتى الحشود الذين يريدون فقط "رؤية ما سيحدث" . (١١) الناس كانوا سهربون من المناطق الحضرية. في النهاية ، سيتم تدمير العديد من الأماكن بسبب التداعيات النووية ، وسيتم التخلي عن المدن ، ولن تكون هناك حاجة للكهرباء أو الزجاج. وحيازة المواد المحظورة ، أو المعدات التي يمكن استخدامها في صنعها ، سيعاقَب عليها بقسوة ، مثل الإعدام الفوري. لتطبيق هذه الأحكام ، ستخضع المجتمعات لمراقبة صارمة – شبكات المخبرين التي تحفزها المكافآت الكبيرة ، واقتحام الشرطة المتكرر للأحياء الخاصة ، والمراقبة الرقمية المستمرة ، وهكذا دواليك.

ذلك هو السيناريو المتفائل. وفي سيناريو أكثر تشاؤمًا ، سينهار القانون والنظام تمامًا وقد تنقسم المجتمعات إلى فصائل تشن حروبًا أهلية بالأسلحة النووية ، مما يؤدي إلى المجاعة والأوبئة. قد ينتهي التفكك فقط عندما يتقلص المجتمع إلى درجة أن لا أحد يستطيع بعدها تجميع قنبلة ومؤخر تفجير من المواد المخزنة أو خردة أطلال المدينة. حتى في ذلك الحين ، فإن الرؤية الخطيرة – بمجرد إثبات أهميتها وبشكل مذهل – سيجري تذكرها وتناقلها عبر الأجيال. إذا بدأت الحضارة في الصعود من تحت الرماد ، فإن المعرفة ستظل في حالة انتظار ، ومستعدة للانقضاض بمجرد أن يتعلم الناس مرة أخرى كيفية صنع الألواح الزجاجية ومولدات التيار الكهربائي. وحتى إذا تم نسيان المعرفة ، فسيتم إعادة اكتشافها بمجرد استئناف أبحاث الفيزياء النووية.

لقد كنا محظوظين بأن أصبح صنع الأسلحة النووية صعبًا.

فرضية العالم المعرض للخطر The vulnerable world hypothesis

نحن نعلم الآن أنه لا يمكن لأحد أن يتوصل إلى انفجار نووي بمجرد استخدام لوح من الزجاج ، ونوع من المعدن ، وبطارية. إن صنع القنبلة الذرية يتطلب عدة كيلوجرامات من المواد الانشطارية التي يصعب إنتاجها. لقد وصلنا إلى كرة رمادية هذه المرة. لكن مع كل عمل من أعمال الاختراع ، نصل إلى داخل الجرة من جديد.

دعونا نقدم الفرضية القائلة بأن جرة الإبداع تحتوي على كرة سوداء واحدة على الأقل. يمكننا الإشارة إلى هذا على أنه فرضية العالم المعرض للخطر (VWH) . بشكل بديهي ، الفرضية هي أن هناك مستوى من التكنولوجيا يتم فيه تدمير الحضارة بشكل شبه مؤكد ما لم يتم تنفيذ درجات غير عادية وغير مسبوقة تاريخياً من السياسة الوقائية و / أو الحوكمة العالمية. وبتعبير أكثر دقة:

أطروحة العالم المعرض للخطر VWH : إذا استمر التطور التكنولوجي ، فسيتم في مرحلة ما تحقيق مجموعة من القدرات التي تجعل تدمير الحضارة أمرًا محتملاً للغاية ، ما لم تَخرج الحضارة بدرجة كافية من الحالة الافتراضية شبه الفوضوية.

وأعني بـ "الحالة الافتراضية شبه الفوضوية" نظامًا عالميًا يتميز بثلاث خصائص: (١٢)

١- محدودية القدرة على العمل السياسي الوقائي. ليس لدى الدول وسائل موثوقة بما فيه الكفاية للمراقبة والاعتراض في الوقت الحالي بحيث يصبح من المستحيل فعليًا على أي فرد أو مجموعة صغيرة داخل أراضيها القيام بأعمال غير قانونية - لا سيما الإجراءات التي لا يحبذها أكثر من ٩٩ % من السكان.

٢- القدرة المحدودة للحوكمة العالمية. لا توجد آلية موثوقة لحل مشكلات التنسيق العالمي وحماية الملكية العامة العالمية - لا سيما في المواقف عالية المخاطر التي تنطوي على أمور حيوية تمس الأمن القومي.

٣- الدوافع المتنوعة. هناك توزيع بشري واسع ومعروف للدوافع يمثله عدد كبير من الجهات الفاعلة (على مستوى الفرد ومستوى الدولة) - على وجه الخصوص ، هناك العديد من الجهات الفاعلة التي تحفزها إلى حد كبير تصورات المصلحة الذاتية (مثل المال ، والسلطة ، والمكانة ، والراحة ، والملاءمة أو الارتياح) وهناك بعض الفاعلين (المتبقي المروع) الذين سيتصرفون بطرق تدمر الحضارة حتى بتكلفة عالية عليهم. (٣)

يمكن تفسير مصطلح "تدمير الحضارة" في التعريف أعلاه بطرق مختلفة ، مما ينتج عنه إصدارات مختلفة من فرضية العالم المعرض للخطر VWH. على سبيل المثال ، يمكن للمرء أن يحدد فرضية العالم المعرض للخطر الوجودي existential-risk vulnerable world hypothesis (x-VWH) ، والتي من شأنها أن تنص على أنه في مستوى معين من التكنولوجيا ، بشكل افتراضي ، تحدث كارثة وجودية ، تنطوي على انقراض الحياة الذكية التي تنشأ من الأرض أو التدمير الدائم لإمكاناتنا المستقبلية لتحقيق القيمة. ومع ذلك ، هنا سوف نخفض الحد الفاصل قليلا. أحد الشواغل الرئيسية في السياق الحالي هو ما إذا ما كانت عواقب استمرار الحضارة في حالة التعثر شبه الفوضوي الحالية تعد عواقب كارثية بما يكفي لكي تفوق الاعتراضات المقبولة على التطورات الجذرية التي قد تكون مطلوبة للخروج من هذه الحالة. إذا كان هذا هو المعيار ، فإن درجة أقل من الانقراض البشري أو الكارثة الوجودية سوف تبدو كافية. على سبيل المثال ، حتى أولئك الذين يشككون بشدة في المراقبة الحكومية سيكون من المفترض أن يفضلوا زيادة كبيرة في مثل هذه المراقبة إذا كانت ضرورية حقًا لمنع الدمار على مستوى المنطقة كلما سنحت الظروف. وبالمثل ، فإن الأفراد الذين يقدرون العيش في دولة ذات

سيادة قد يفضلون بدرجة معقولة العيش في ظل حكومة عالمية نظرًا لافتراض أن البديل قد ينطوي على شيء رهيب يشبه المحرقة النووية. لذلك ، فإننا نقرر أن مصطلح (تدمير الحضارة) في فرضية عالم معرض للخطر VWH يشير (ما لم يتم تحديد مغاير لذلك) إلى أي حدث مدمر على الأقل بمثل خطورة وفاة ١٥ في المائة من سكان العالم أو انخفاض إجمالي الناتج المحلي العالمي بنسبة أكثر من ٥٠ في المائة لمدة تزيد عن عقد من الزمن. (١٣)

ليس الغرض الأساسي من هذه الورقة القول بأن فرضية عالم معرض للخطر VWH صحيحة. (إنني أعتبرها سؤالًا مفتوحًا ، على الرغم من أنها قد تبدو لي غير معقولة ، بالنظر إلى الأدلة المتاحة ، فقط رجاء الوصول إلى الثقة التامة بأن VWH خاطئة.) بدلاً من ذلك ، فإن المساهمة الرئيسية المزعومة هنا هي أن VWH ، إلى جانب المفاهيم والتفسيرات ذات الصلة ، مفيدة في مساعدتنا لإبراز اعتبارات وإمكانيات مهمة فيما يتعلق بالوضع الاستراتيجي الكلي للبشرية. لكن هذه الاعتبارات والإمكانيات تحتاج إلى مزيد من التحليل ، ودمجها مع اعتبارات أخرى تقع خارج نطاق هذه الورقة ، قبل أن تتمكن من تقديم أية استنتاجات سياسية نهائية.

بعض التوضيحات الإضافية قبل المضي قدمًا. تستخدم هذه الورقة كلمة "التكنولوجيا" بمعناها الواسع. وبالتالي ، من حيث المبدأ ، لا نضع في اعتبارنا فقط الآلات والأجهزة المادية ولكن أيضًا الأنواع الأخرى من القوالب والإجراءات الفعالة من الناحية العملية – بما في ذلك الأفكار العلمية ، والتصميمات المؤسسية ، والتقنيات التنظيمية ، والأيديولوجيات ، والمفاهيم ، والميمات (**) memes – باعتبارها تشكل كرات سوداء تكنولوجية محتملة. (١٤)

يمكننا التحدث عن ثغرات نقاط الخطورة المحتملة وعن إغلاقها. في سيناريو "الأسلحة النووية السهلة" ، تبدأ فترة نقطة الخطورة المحتملة عندما يتم اكتشاف الطريقة السهلة لإنتاج التفجيرات النووية. وتنتهي عندما يتم الوصول إلى مستوى معين من التكنولوجيا يجعل من الممكن بشكل معقول منع التفجيرات النووية من التسبب في أضرار غير مقبولة – أو يجعل من غير الممكن إنتاج تفجيرات نووية (بسبب التراجع التكنولوجي) . (١٥) إذا لم تكن هناك تقنية وقائية ممكنة (كما في حالة الأسلحة النووية ، على سبيل المثال ، قد لا تكون كذلك) والتراجع التكنولوجي لا يحدث ، وعندئذٍ يصبح العالم عرضة للخطر بشكل دائم.

يمكننا أيضًا التحدث عن استقرار العالم (فيما يتعلق ببعض نقاط الخطورة المحتملة) إذا تم الخروج من حالة التقصير شبه الفوضوي بطريقة تمنع نقاط الخطورة المحتملة من التسبب في كارثة فعلية. تعتمد الطرق ، التي يجب أن يتم بها تغيير حالة التقصير شبه الفوضوي من أجل تحقيق الاستقرار ، على تفاصيل نقاط الضعف المعنية. في قسم لاحق ، سوف نناقش الوسائل الممكنة التي يمكن من خلالها تحقيق الاستقرار في العالم. في الوقت الحالي ، نلاحظ ببساطة أن فرضية العالم المعرض للخطر VWH لا تعني أن الحضارة محكوم عليها بالفناء.

تصنيف نقاط الخطورة المحتملة **Typology of vulnerabilities**

يمكننا تحديد أربعة أنواع من نقاط الخطورة الحضارية المحتملة:

النوع ١ (الأسلحة النووية السهلة)

النوع الأول ، كما هو الحال في سيناريو "الأسلحة النووية السهلة" ، يصبح من السهل جدًا على الأفراد أو المجموعات الصغيرة إحداث دمار شامل:

الخطورة المحتملة من النوع الأول: هناك بعض التكنولوجيا المدمرة للغاية وسهلة الاستخدام ، نظرًا للظروف الافتراضية شبه الفوضوية ، فإن تصرفات الجهات الفاعلة في المخلفات المروعة تجعل الدمار الحضاري محتملاً للغاية.

لاحظ أنه عند تحديد ما إذا كان السيناريو يمثل نقطة الخطورة المحتملة من النوع ١ ، توجد علاقة عكسية بين السهولة التي يمكن من خلالها التسبب في وقوع حادث وبين القدرة التدميرية للحادث. فكلما زاد التدمير الناتج عن حادثة واحدة ، قلت سهولة التسبب في مثل هذا الحادث حتى نتمكن من تشخيص وجود نقطة خطورة محتملة من النوع الأول.

وبالتالي ، ضع في اعتبارك سيناريو "الأسلحة النووية السهلة للغاية" ، حيث يمكن لأي شخص نصف ذكي أن يصنع سلاحًا حراريًا نوويًا بسهولة في حوض المطبخ في فترة ما بعد الظهيرة: هذا سوف يعد بالتأكيد خطورة حضارية محتملة. بمقارنة هذا مع سيناريو (الأسلحة النووية السهلة إلى حد ما) ، حيث يتطلب الأمر فريقًا مكونًا من خمسة أفراد شبه المهرة يكدحون لمدة عام كامل لإنتاج جهاز ضخم بوزن بضعة أطنان: قد لا يرقى هذا إلى مستوى نقطة خطورة

حضارية محتملة عالية التأثير. يبدو أنه من الممكن ، في سيناريو (الأسلحة النووية السهلة إلى حد ما) ، أن الغالبية العظمى من المدن تنجو من الدمار ، على الرغم من أن التهديد الذي تشكله منظمة إرهابية تتمتع بموارد جيدة ، مثل أوم شينريكيو Aum Shinrikyo عام ١٩٩٥ أو القاعدة Al-Qaeda عام ٢٠٠١ ، سوف يزداد بصورة جوهرية. ومع ذلك ، ضع في اعتبارك سيناريو آخر ، (الموت البيولوجي السهل إلى حد ما) ، والذي يتطلب مرة أخرى فريقًا نصف ماهر مكون من خمسة أشخاص يعمل لمدة عام لتطبيق تقنية الكرة السوداء ، باستثناء أنها هذه المرة فاعل بيولوجي ، إطلاق نقطة واحدة منه تكفي لقتل المليارات. في "الموت البيولوجي السهل بشكل معتدل" ، سيتم الوصول إلى عتبة نقطة الخطورة المحتملة من النوع الأول. إذا تطلب تدمير الحضارة فقط أن تنجح مجموعة واحدة في مهمة على مستوى سهل إلى حد ما ، فمن المحتمل أن يتم تدمير الحضارة في غضون بضع سنوات في حالة افتراضية شبه فوضوية. في الواقع ، لقد سعت كل من أوم شينريكيو والقاعدة للحصول على أسلحة نووية وبيولوجية ، ومن المرجح أنهما اختارتا استخدامها (انظر على سبيل المثال Danzig et al., 2011 ؛ Olson, 1999 ؛ Mowatt-Larssen and Allison, 2010) .

لذلك توجد الخطورة المحتملة من النوع ١ إذا كان من السهل للغاية التسبب في قدر متوسط من الضرر أو كان من السهل إلى حد ما التسبب في قدر كبير من الضرر. (١٦) السبب في أن تقنية الكرة السوداء التي تتيح مقادير معتدلة فقط من الضرر لكل حادث يمكن اعتبارها نقطة خطورة محتملة من النوع الأول هي أنه – إذا كانت التكنولوجيا سهلة الاستخدام بدرجة كافية – من شبه المؤكد حدوث عدد كبير من هذه الحوادث. خذ السيناريو الذي يسهل على الفرد العادي صنع قنبلة هيدروجينية تخترق المدينة، فهذا ليس بالضرورة سيناريو يمكن فيه لفرد واحد تدمير الحضارة. سيظل تصنيع مئات القنابل ونقلها إلى مئات المدن دون أن يتم القبض عليها مسعىً مهولاً حتى لو كان صنع قنبلة واحدة أمرًا سهلاً إلى حد ما. ومع ذلك ، فإن سيناريو "الأسلحة النووية السهلة" يمثل نقطة خطورة حضارية محتملة لأنه من المعقول أنه سيكون هناك في الواقع مئات الأفراد الذين سيدمر كل منهم مدينة واحدة على الأقل في ظل تلك الظروف.

إن هذا ينبع تقريبًا من قانون الأعداد الكبيرة جنبًا إلى جنب مع الافتراض المعقول أنه بالنسبة لأي شخص يتم اختياره عشوائيًا ، هناك فرصة صغيرة لأن يكون لديه الدافع لإحداث هذا النوع من الدمار – سواء كان ذلك بسبب الكراهية الأيديولوجية ، أو العدمية التدميرية ، أو الانتقام من الظلم الذي يتصوره ، أو كجزء من مؤامرة ابتزاز ، أو بسبب الأوهام أو المرض العقلي ، أو ربما لمجرد رؤية ما سيحدث. وبالنظر إلى تنوع الشخصية والظروف البشرية ، بالنسبة لأي عمل غير حكيم ، أو غير أخلاقي ، أو يهدد الذات ، هناك جزء متبقي من البشر الذين سوف يختارون اتخاذ هذا الفعل. هذا معقول بشكل خاص إذا كان الفعل المعني يمثل تكلفة ملحوظة ثقافيًا – كما هو الحال في كل مكان بعد حدوث هجوم نووي. بعبارة أخرى ، "الأسلحة النووية السهلة" هي توضيح لعالم معرض للخطر لأنه يبدو أن المخلفات المروعة لها تقاطع كبير بما يكفي مع مجموعة الجهات الفاعلة التي تتمتع بالسلطة التي يتوقع المرء أن ينتج عنها قدر مدمر من تدمير الحضارة.

النوع ٢ أ (الضربة الأولى الآمنة)

إن التكنولوجيا التي "تضفي الطابع الديمقراطي" على الدمار الشامل ليست النوع الوحيد من الكرة السوداء التي يمكن إخراجها من الجرة. نوع آخر هو التكنولوجيا التي تحفز بقوة الجهات الفاعلة القوية على استخدام قوتها لإحداث دمار شامل. مرة أخرى يمكننا أن ننتقل إلى التاريخ النووي للتوضيح.

بعد اختراع القنبلة الذرية والاحتكار النووي الأمريكي قصير المدى ، تبع ذلك سباق تسلح بين الولايات المتحدة والاتحاد السوفيتي. جمعت القوى العظمى المتنافسة ترسانات مذهلة ، تجاوزت ٧٠٠٠٠ رأس نووي عام ١٩٨٦ ، أكثر مما يكفي لتدمير الحضارة (Norris and Kristensen, 2010). وفي حين يبدو أن الوعي العام بمخاطر الحرب الباردة قد تلاشى منذ نهايتها السلمية عام ١٩٩١ ، إلا أن المجتمع الأكاديمي – مستفيدًا من فتح الأرشيفات السرية السابقة وشهادات صانعي السياسات والضباط والمحللين المتقاعدين – كشف النقاب عن مجموعة مقلقة من الممارسات والحوادث التي يبدو أنها دفعت العالم مرارًا وتكرارًا إلى حافة الهاوية. (١٧) وقد قال بعض العلماء أنه بفضل قدر كبير من الحظ تم تجنب محرقة نووية. (١٨)

وسواء كانت النجاة من الحرب الباردة تتطلب الكثير من الحظ أو القليل فقط ، يمكننا بسهولة تخيل واقعي مضاد تكون فيه احتمالات تجنب حريق نووي أسوأ بكثير. هذا صحيح حتى لو افترضنا أن الأسلحة النووية لا يمكن إنتاجها إلا من قبل الدول الكبيرة المتقدمة تقنيًا (وبالتالي تمييز هذه الحالة عن نقطة الضعف من النوع ١ من "الأسلحة النووية السهلة"). يمكن للواقع المضاد أن ينطوي على تغييرات في حدود الاحتمالات التكنولوجية التي من شأنها أن تجعل سباق التسلح أقل استقرارًا.

على سبيل المثال ، يُعتقد على نطاق واسع بين الاستراتيجيين النوويين أن تطوير قدرة الضربة الثانية الآمنة بشكل معقول من قِبل كلتا القوتين العظميين بحلول منتصف الستينيات قد خلق ظروفًا تتيح "الاستقرار الاستراتيجي" (Colby and Gerson, 2013) . قبل هذه الفترة ، عكست الخطط الحربية الأمريكية ميلًا أكبر بكثير ، في أية حالة أزمة ، نحو شن ضربة نووية استباقية ضد الترسانة النووية للاتحاد السوفيتي. كان يُعتقد أن إدخال صواريخ باليستية عابرة للقارات تعتمد على الغواصات النووية مفيد بشكل خاص لضمان قدرات الضربة الثانية (وبالتالي "التدمير المتبادل المؤكد") لأنه كان يُعتقد على نطاق واسع أنه من المستحيل عمليًا على المعتدي القضاء على أسطول العدو في الهجوم الأوَّلي. (١٩) يمكن أيضًا استخدام الاستراتيجيات الأخرى لضمان القدرة على الضربة الثانية ، ولكنها تنطوي على عيوب؛ على سبيل المثال ، كان أحد الخيارات ، الذي استخدمته الولايات المتحدة لفترة وجيزة ، هو امتلاك مجموعة من القاذفات النووية بعيدة المدى في حالة تأهب جوي مستمر (ساجان ,Sagan 1995) . كان هذا البرنامج مكلفًا للغاية وزاد من مخاطر الهجمات العرضية أو غير المصرح بها. كان الخيار الآخر هو بناء صوامع صواريخ أرضية صلبة: بأعداد كافية ، يمكنها من حيث المبدأ توفير ضمان قدرة الضربة الثانية على جانب واحد ؛ ومع ذلك ، فإن مثل هذه الترسانة الكبيرة من شأنها أن تهدد بتوفير القدرة على توجيه ضربة أولى آمنة ضد الجانب الآخر ، وبالتالي زعزعة استقرار أية أزمة مرة أخرى. قاذفات الصواريخ المحمولة البرية، والتي يصعب مهاجمتها أكثر من الصواريخ التي تعتمد على الصوامع ، وفرت في النهاية بعض الاستقرار عندما تم نشرها من قبل الاتحاد السوفيتي في عام ١٩٨٥ ، قبل بضع سنوات من نهاية الحرب الحرب الباردة (Brower, 1989) .

لذلك نضع في اعتبارنا الواقع المضاد الذي تكون فيه القوة المضادة الوقائية ضد تلقي الضربة أكثر جدوى. تخيل نوعا من التقنيات يجعل من السهل تتبع غواصات الصواريخ الباليستية. يمكننا أيضًا أن نتخيل لو أن الأسلحة النووية كانت أكثر هشاشة ، لذا فإن نصف القطر الذي سيتم تدمير سلاح نووي داخله بتفجير سلاح نووي آخر كان أكبر بكثير مما هو عليه في الواقع. (٢٠) في ظل هذه الظروف ، قد يكون من المستحيل ضمان قدرة ضربة ثانية. لنفترض ، علاوة على ذلك ، أن التكنولوجيا كانت بكيفية تجعل من الصعب للغاية اكتشاف عمليات إطلاق الصواريخ البالستية، مما يجعل استراتيجية الإنذار عند الإطلاق غير قابلة للتطبيق تمامًا. كان من الممكن بعد ذلك تضخيم عدم استقرار أزمة الحرب الباردة إلى حد كبير. أيًا كان الجانب الذي سيُضرب أولاً ، فإنه سينجو دون أن يصاب بأذى نسبيًا (أو ربما كان يعتقد على الأقل أنه سينجو ، لأن احتمالية حدوث شتاء نووي (تلقي ضربة نووية) تم تجاهلها إلى حد كبير من قِبل مخططي الحرب في ذلك الوقت) (Badash, 2001; Ellsberg, 2017) . (٢١) الجانب الأقل عدوانية سوف يدمَّر تماما. في مثل هذه الحالة ، يمكن أن يؤدي الخوف المتبادل بسهولة إلى الاندفاع نحو حرب شاملة (Schelling, 1960) .

يمكن أن تؤدي تغييرات المعيار التكنولوجي الأخرى إلى زيادة احتمالية الهجمات. في العالم الحقيقي ، يكون عامل "الجذب" الرئيسي في الضربة النووية الأولى هو أنها ستخفف من الخوف من أن يصبح المرء ضحية مثل هذه الضربة ؛ لكن يمكننا أن نتخيل واقعًا مضادًا حيث توجد أيضًا فوائد للعدوان النووي ، تتجاوز إزالة السلبية. من المفترض أنه كان من الممكن بطريقة ما جني مكاسب اقتصادية كبيرة من البدء بعدوان نووي واسع المدى. (٢٢) قد يكون من الصعب فَهم كيف يمكن أن يكون هذا هو الحال ، ومع ذلك يمكن للمرء أن يتخيل بعض تقنيات التصنيع الآلي أو تكنولوجيا الطاقة التي تجعل الموارد المادية أكثر قيمة ؛ أو يمكن للنمو السكاني المدعوم بالتكنولوجيا أن يجعل الأراضي الزراعية مورداً أكثر حيوية (Drexler,1986). يعتقد بعض دارسي العلاقات الدولية أن صافي الأرباح الاقتصادية للغزو قد انخفض بشكل كبير في مرحلة ما بعد الصناعة وأن هذا التراجع كان مساهماً رئيسياً في السلام. (٢٣) إذا تمت إضافة دوافع اقتصادية وطنية قوية مرة أخرى إلى أسباب الحرب الأخرى (مثل الاهتمام بأمن الفرد ، والنزاعات حول القيم غير الاقتصادية ، والحفاظ على السمعة

الوطنية ، وتأثير مجموعات المصالح العدوانية الخاصة بشكل خاص ، من بين أمور أخرى) قد تصبح النزاعات المسلحة أكثر شيوعًا والحرب النووية واسعة النطاق أكثر احتمالا.

في هذه الأمثلة ، لا تنشأ قابلية التأثر بالخطر من زيادة سهولة التدمير ، ولكن من الأفعال التي تؤدي إلى التدمير التي تدعمها حوافز أقوى. سوف نسمي ثغرات الخطورة المحتملة هذه النوع-٢ . على وجه التحديد ، سيناريو مثل "الضربة الأولى الآمنة" ، حيث يتم تحفيز فعل معين مدمر للغاية ، يجب أن نشير إليه باعتباره النوع ٢ أ :

الخطورة المحتملة من النوع ٢ أ : يوجد مستوى معين من التكنولوجيا يتمتع فيه الفاعلون الأقوياء بالقدرة على إحداث أضرار مدمرة للحضارة ، وفي حالة التقصير شبه الفوضوي ، يواجهون دوافع استخدام هذه القدرة.

سوف نرى بعض الأمثلة الأخرى للخطورة المحتملة من النوع ٢ أ في ما يلي ، حيث تتخذ "الأضرار المدمرة للحضارة" شكل عوامل خطورة خارجية.

النوع ٢ ب ("الاحتباس الحراري العالمي الأسوأ")

هناك طريقة أخرى يمكن أن يكون العالم عرضة للخطر من خلالها ؛ هذه الطريقة يمكننا توضيحها بواقعية معاكسة تتعلق بتغير المناخ.

في العالم الحقيقي ، نلاحظ ارتفاعًا ملحوظًا في متوسط درجة الحرارة العالمية ، يُعتقد على نطاق واسع أن أسبابها الأساسية هي انبعاثات الغازات الدفيئة التي يسببها الإنسان مثل ثاني أكسيد الكربون والميثان وأكسيد النيتروز (Stocker et al., 2014) . وتتنوع الاعتراضات اعتمادا على سيناريو الانبعاثات وافتراضات نماذج الحالات ، ولكن التنبؤات التي تشير إلى ارتفاع في متوسط درجة الحرارة يتراوح بين ٣ درجات مئوية و ٤,٥ درجة مئوية في ٢١٠٠ (مقارنة بعام ٢٠٠٠) ، وذلك في ظل عدم وجود أي إجراء مهم لتقليل الانبعاثات ، فهذا أمر نموذجي تمامًا ، انظر ((Stocker et al. (2014, table 12.2) . من المتوقع أن تكون تأثيرات هذا الاحترار – على مستويات سطح البحر ، وأنماط الطقس ، والنظم البيئية ، والزراعة – سلبية على رفاهية الإنسان ، انظر ((Field et al. (2014, figure 10-1) . تنبعث الغازات الدفيئة من مجموعة واسعة من الأنشطة ، بما في ذلك في الصناعة ، والنقل ، والزراعة،

وإنتاج الكهرباء ، ومن جميع أنحاء العالم ، خاصةً من الدول الصناعية أو التي تشق طريقها في التصنيع. لقد فشلت الجهود المبذولة للحد من الانبعاثات حتى الآن في تحقيق تأثير عالمي كبير (Friedlingstein et al., 2014) .

الآن ، يمكننا أن نتخيل وضعًا تكون فيه مشكلة الاحتباس الحراري أشد خطورة مما تبدو عليه في الواقع. على سبيل المثال ، يمكن أن تكون حساسية المناخ العابرة (مقياسا لتغير متوسط المدى في متوسط درجة حرارة سطح الأرض عالميا والتي تنتج عن نوع من التأثير ، مثل مضاعفة ثاني أكسيد الكربون في الغلاف الجوي) ومن المحتمل أن يصبح أكبر بكثير مما هو عليه الآن (Shindell, 2014) . إذا كان أكبر عدة مرات من قيمته الحقيقية ، كنا سنواجه ارتفاعًا في درجة الحرارة ، على سبيل المثال ، ١٥ درجة أو ٢٠ درجة مئوية بدلاً من ٣ درجات مئوية – احتمال وجود إمكانية لتدمير الحضارة أكبر بكثير من التوقعات الفعلية. (٢٤)

يمكننا أيضًا أن نتخيل انحرافات أخرى عن الواقع من شأنها أن تجعل من الاحتباس الحراري العالمي مشكلة أسوأ. إن أنواع الوقود الأحفوري كان بالإمكان جعلها أكثر وفرة مما هي عليه ، وأن تكون متاحة في صورة رواسب قابلة للاستغلال بتكلفة أقل ، الأمر الذي كان سيشجع زيادة استهلاك بدرجة أكبر. في نفس الوقت ، كان من الممكن ارتفاع تكلفة بدائل الطاقة النظيفة وأن تكون أكثر تحديًا من الناحية التكنولوجية. وكان من المحتمل أن يكون الاحتباس الحراري مشكلة أسوأ إذا توفرت حلقات تقييم إيجابية أقوى وغير خطية ، مثل المرحلة الأولية التي يتم فيها تحميل الغلاف الجوي تدريجيًا بغازات دفيئة دون تأثير ملحوظ أو ضار ، تليها مرحلة ثانية ترتفع فيها درجات الحرارة بشكل مفاجئ. وللحصول على تهديد حضاري حقيقي من الاحتباس الحراري العالمي ، قد يكون من الضروري أيضًا النص ، بشكل معاكس ، على أن التخفيف من خلال الهندسة الجيولوجية غير ممكن.

الخطورة المحتملة التي جرى توضيحها باستخدام سيناريو "الاحتباس الحراري العالمي الأسوأ" تختلف عن سيناريو خطورة النوع ٢ أ مثل "الضربة الأولى الآمنة". في الخطورة من النوع ٢ أ يمتلك بعض الفاعلين القدرة على اتخاذ بعض الإجراءات – مثل شن ضربة نووية أولية – وهذا مدمر بدرجة كافية لتدمير الحضارة. في سيناريو "الاحتباس الحراري العالمي الأسوأ" ، لا توجد حاجة لوجود مثل هذه الفاعلية. بدلاً من ذلك ، فيما نسميه الخطورة المحتملة من النوع ٢

ب ، هناك عدد كبير من الفاعلين غير المهمين بشكل فردي الذين يتم تحفيز كل منهم (في الحالة الافتراضية شبه الفوضوية) لاتخاذ بعض الإجراءات التي تساهم بشكل طفيف فيما يصبح بطريقة تراكمية مشكلة تدمير حضارية:

الخطورة المحتملة من النوع ٢ ب: هناك مستوى من التكنولوجيا يواجه فيه عدد كبير من الفاعلين ، في حالة افتراضية شبه فوضوية ، ضغوطاً لاتخاذ بعض الإجراءات المدمرة بشكل طفيف بحيث يكون التأثير المركب من تلك الإجراءات تدميرا حضاريا.

العامل المشترك بين النوع ٢ أ والنوع ٢ ب هو أنه في كلتا الحالتين ، يواجه الفاعلون الذين لديهم القدرة على إلحاق الضرر ضغوطاً من شأنها أن تشجع مجموعة واسعة من الجهات الفاعلة ذات الدوافع الطبيعية في وضعهم على مواصلة مسار الفعل الذي يؤدي إلى الضرر. لن يكون الاحترار العالمي مشكلةً عندما تختار فئة صغيرة فقط من أولئك الفاعلين الذين يمكنهم قيادة السيارات أو قطع بعض الأشجار القيام بذلك ؛ ولكن تنشأ المشكلة فقط لأن الكثير من الفاعلين يتخذون هذه الخيارات. ولكي يتخذ الكثير من الجهات الفاعلة هذه الاختيارات ، يجب أن تكون الاختيارات مدعومة بحوافز تتمتع بجاذبية واسعة (مثل المال ، والوضع ، والتوافق). وبالمثل ، إذا اختارت واحدة فقط من بين كل مليون جهة فاعلة يمكنها شن ضربة نووية أولى القيام بذلك ، فلن يكون الأمر مقلقًا للغاية إذا كان هناك عدد قليل من الجهات الفاعلة التي تمتلك تلك القدرة ؛ لكن الأمر يثير القلق إذا كان إطلاق ضربة نووية مدعومًا بقوة بالحوافز التي تروق للجهات الفاعلة ذات الدوافع الطبيعية (مثل الدافع وراء استباق ضربة من قبل الخصم). هذا على النقيض من الخطورة المحتملة من النوع ١ ، حيث تنشأ المشكلة من الانتشار الواسع النطاق للقدرة التدميرية. فقط جهة فاعلة واحدة لديها قيم غير عادية سوف تختار ، بتكلفة باهظة ومخاطرة بنفسها ، تفجير مدينة أو إطلاق العنان لمرض كاسح ؛ المشكلة في هذه الحالة هي أنه إذا كان لدى العديد من الجهات الفاعلة مثل هذه القدرة ، فإن مجموعة كبيرة منها لديها أيضًا دوافع تنبؤية لن تنعدم لديهم هذه القدرة.

النوع صفر (حالات الخنق المفاجئ)

في عام ١٩٤٢ ، خطر ببال إدوارد تيلر Edward Teller ، أحد علماء مانهاتن ، أن الانفجار النووي من شأنه أن يخلق درجة حرارة غير مسبوقة في تاريخ الأرض ، مما ينتج عنه ظروف مماثلة لتلك الموجودة في مركز الشمس ، وأن هذا يمكن أن يؤدي إلى ما يمكن تصوره تفاعلا حراريا نوويا مستداما بطريقة ذاتية في الهواء أو الماء المحيط (Rhodes, 1986) . لقد أدرك روبرت أوبنهايمر Robert Oppenheimer ، رئيس مختبر لوس ألاموس Los Alamos ، أهمية مخاوف تيلر على الفور. أبلغ أوبنهايمر رئيسه وأمر بإجراء مزيد من التحليلات للتحقيق في هذا الاحتمال. أشارت هذه التحليلات إلى أن الاشتعال في الغلاف الجوي لن يحدث. ثم تم تأكيد هذا التوقع في عام ١٩٤٥ من خلال اختبار ترينيتي Trinity ، والذي تضمن تفجير أول قنبلة نووية في العالم. (٢٥)

في عام ١٩٥٤ ، أجرت الولايات المتحدة اختبارًا نوويًا آخر ، وهو اختبار قلعة برافو Castle Bravo ، والذي تم التخطيط له كتجربة سرية مع تصميم مبكر لقنبلة نووية حرارية تعتمد على الليثيوم. يحتوي الليثيوم ، مثل اليورانيوم ، على اثنين من النظائر المهمة: ليثيوم ٦ وليثيوم ٧ . بعد الاختبار ، قدَّر العلماء النوويون الناتج بأنه ٦ ميجا طن (مع مدى غير مؤكد من ٤ إلى ٨ ميجا طن). لقد افترضوا أن الليثيوم ٦ وحده هو الذي يساهم في التفاعل ، لكنهم كانوا خاطئين. لقد ساهم الليثيوم ٧ في إنتاج طاقة أكثر من الليثيوم ٦ ، وانفجرت القنبلة بقوة ١٥ ميغا طن – أكثر من ضعف ما قدَّروه (وما يعادل حوالي ١٠٠٠ مرة من قوة انفجار هيروشيما) . لقد دمر الانفجار القوي بشكل غير متوقع الكثير من معدات الاختبار ، كما تسبب الغبار الإشعاعي المتساقط في تسمم سكان الجزر الواقعة في مسار الريح وطاقم قارب صيد ياباني ، مما تسبب في وقوع حادث دولي.

قد نعتبر أنه من حسن الحظ أن تقدير قلعة برافو Castle Bravo كان غير صحيح ، وتقدير ما إذا كان اختبار ترينيتي Trinity لن يشعل الغلاف الجوي. في المقابل ، إذا كان الغلاف الجوي عرضة للاشتعال عن طريق تفجير نووي ، وإذا كان من السهل نسبيًا التغاضي عن هذه الحقيقة – دعونا نقول أنه بنفس السهولة التي جرى بها التغاضي عن مساهمة الليثيوم ٧ في اختبار قلعة برافو Castle Bravo – ثم تنتهي مسيرة البشر (ومسيرة كل الحياة الأرضية)

في عام ١٩٤٥. يمكننا تسمية هذا السيناريو "قلعة برافيشيمو Castle Bravissimo" (إسم مركب من برافو و هيروشيما – المترجم) .

عندما نسحب كرة من جرة/جعبة الاختراع ، يمكننا تصور احتمال حدوث دمار عرضي. عادة ، تعد هذه الخطورة المحتملة ضئيلة ؛ ولكن في بعض الحالات قد تكون مهمة ، لا سيما عندما تولِّد التكنولوجيا المعنية نوعًا من الاضطراب الجديد في الطبيعة أو تنتج ظروفًا غير مسبوقة تاريخيًا. يشير هذا إلى أنه يجب علينا أن نضيف إلى تصنيفنا فئة أخرى ، فئة الدمار الحضاري العرَضي بسبب التكنولوجيا:

الخطورة المحتملة من النوع صفر: هناك بعض التقنيات التي تحمل خطورة خفية حيث تكون نتيجتها الافتراضية عند اكتشافها هي تدمير حضاري غير مقصود. (٢٦)

من المفيد أن نلاحظ ، مع ذلك ، أن "قلعة برافيشيمو" ليست توضيحًا مثاليًا لخطورة النوع صفر. لنفترض أن الحسابات الدقيقة أظهرت أن هناك احتمالًا بنسبة ١% أن يؤدي انفجار نووي إلى إشعال الغلاف الجوي والمحيطات وبالتالي انتهاء الحياة على الأرض. افترض ، علاوة على ذلك ، أنه كان من المعروف أنه لحل المشكلة بشكل أكبر وإثبات أن الفرصة كانت صفراً (أو بدلاً من ذلك ، كانت الفرصة درجة واحدة) سوف يستغرق الأمر ١٠ سنوات أخرى من الدراسة الدقيقة. من غير الواضح ، في ظل هذه الظروف ، ما الذي كان سيقرره قادة مشروع مانهاتن. كان من المفترض أن يعتقدوا أنه من المرغوب فيه بشدة أن تتوقف البشرية عن تطوير الأسلحة النووية لمدة ١٠ سنوات أخرى على الأقل. (٢٧) من ناحية أخرى ، كانوا سوف يخشون أن يكون لدى ألمانيا مشروع قنبلة متقدم وأن هتلر ربما لن يصدر أمر التوقف. بسبب خطورة تدمير العالم بنسبة ١% . (٢٨) ربما استنتجوا أن خطورة اختبار قنبلة نووية تستحق المخاطرة من أجل تقليل احتمالية أن ينتهي الأمر بألمانيا النازية بالاحتكار النووي.

في هذا الإصدار من "قلعة برافيشيمو" ، يجري تفجير الحضارة عن طريق الصدفة: لم يحاول أحد تدبير حدث مدمر. ومع ذلك ، كانت الجهات الفاعلة الرئيسية عالقة في وضع استراتيجي حفزها على المضي قُدما على الرغم من المخاطر. في هذا الصدد ، يعتبر السيناريو خطورة محتملة من النوع ٢ أ ؛ الضرر المدمر للحضارة هو مجرد احتمال فقط. عندما تصبح

التكنولوجيا النووية ممكنة ، يواجه الفاعلون الأقوياء دوافع ، في حالة افتراضية شبه فوضوية ، لاستخدام هذه التكنولوجيا بطرق تنتج أضرارًا مدمرة للحضارة (والتي تأخذ هنا شكل مخاطر ذات عوامل خارجية) . (٢٩)

وفقًا لذلك ، من أجل تشخيص الخطورة المحتملة من النوع صفر ، فإننا نطلب استيفاء شرط أقوى من مجرد عدم نية الجهات الفاعلة الرئيسية القيام بالتدمير. إننا نشترط أن تعني كلمة "غير مقصود" هنا أن النتيجة العكسية نشأت عن سوء الحظ ، وليست ناشئة عن فشل التنسيق. في الخطورة من النوع صفر ، تقرر الجهات الفاعلة الرئيسية المضي قُدما باستخدام التكنولوجيا ، حتى لو تم تنسيقها بشكل كافٍ ، اعتقادًا منهم بأن الفوائد سوف تفوق التكاليف – لكنهم سوف يتضح أنهم خاطئين ، وأن التكاليف سوف تكون أكبر مما كان متوقعًا ، بما يكفي لإحداث دمار حضاري. (٣٠)

منذ "قلعة برافيشيمو" يجري التزام هذا المعيار بشكل غامض فقط (من غير الواضح في الواقع المضاد الأصلي إلى أي مدى كانت الكارثة ستنجم عن فشل التنسيق وإلى أي مدى بسبب سوء التقدير / سوء الحظ) ، قد يكون من المفيد تقديم مثال أكثر وضوحا للخطورة المحتملة من النوع صفر. وبالتالي ، نفكر في سيناريو "الخنق المفاجئ" الذي تظهر فيه أن بعض التجارب الحديثة في فيزياء الطاقة العالية بدأت عملية تحفيز ذاتي يتم فيها تحويل المادة العادية إلى مادة غريبة ، مما يؤدي إلى تدمير كوكبنا. تم تحليل هذا السيناريو ، وتنويعاته في الأدبيات ، حيث تولِّد تجارب المسرعات / المحفزات ثقوبًا سوداء ثابتة أو تؤدي إلى تحلل حالة الفراغ الاستقراري الهش (Jaffe et al., 2000; Tegmark and Bostrom, 2005) . مثل هذه النتائج سوف تكون مفاجئة للغاية بالفعل ، لأن التحليل يشير إلى أن فرصة حدوثها ضئيلة تمامًا. بالطبع ، مع وجود حظ سيء بما فيه الكفاية ، يمكن أن يقع حدث بطريق الإهمال. ولكن بدلاً من ذلك (والأرجح في هذه الحالة) ، يمكن أن ينطوي التحليل على عيب خفي ، كما فعلت تقديرات Castle Bravo ؛ في هذه الحالة ، قد لا تكون الفرصة ضئيلة للغاية على الإطلاق (Ord et al., 2010) . (٣١)

تحقيق الاستقرار Achieving stabilization

قد تكون حقيقة فرضية العالم المعرض للخطر VWH هي كونها أخبارًا سيئة. لكن ذلك لا يعني أن الحضارة سوف يتم تدميرها. من حيث المبدأ على الأقل ، هناك الكثير من الاستجابات التي يمكن أن تحقق الاستقرار في العالم حتى لو كانت هناك عرضة محتملة للخطر. تذكَّر أننا حددنا الفرضية من حيث تكنولوجيا الكرة السوداء ، مما يجعل الدمار الحضاري على الأرجح مشروطًا باستمرار التطور التكنولوجي واستمرار الحالة الافتراضية شبه الفوضوية. وبالتالي يمكننا نظريًا النظر في الإمكانات التالية من أجل تحقيق الاستقرار:

١- تقييد التطور التكنولوجي.

٢- التأكد من عدم وجود عدد كبير من السكان ينتمون إلى الجهات الفاعلة التي تمثل توزيعًا بشريًا واسعًا وبشكل ملحوظ للدوافع.

٣- إنشاء شرطة وقائية فعالة للغاية.

٤- إنشاء حوكمة عالمية فعالة.

سوف نناقش (٣) و (٤) في الأقسام اللاحقة. هنا ننظر في (١) و (٢) ، وسنوضح أن هاتين النقطتين تحملان وعودًا محدودة فقط كطرق للحماية من مصادر الخطورة الحضارية المحتملة.

التنازل التكنولوجي Technological relinquishment

في شكله العام ، يبدو التنازل التكنولوجي غير واعد للغاية. تذكَّر أننا فسرنا كلمة "تكنولوجيا" على نطاق واسع ؛ بحيث يتطلب التوقف التام للتطور التكنولوجي شيئًا قريبًا من توقف النشاط الابتكاري في كل مكان في العالم. هذا واقعي إلى حد ما ؛ فإذا كان من الممكن القيام بذلك ، فسيكون مكلفًا للغاية – إلى حد تشكيل كارثة وجودية بمعنى الكلمة (أي "الركود الدائم" (Bostrom, 2013) .

على أي حال ، فإن عدم اعتبار التخلي العام عن البحث العلمي والتكنولوجي بدايةً لا يعني أن القيود المحدودة على الأنشطة الابتكارية لا يمكن أن تكون فكرة جيدة. قد يكون من المنطقي التخلي عن اتجاهات التقدم الخطيرة بشكل خاص. على سبيل المثال ، مع استدعاء سيناريو "الأسلحة النووية السهلة" ، سيكون من المعقول عدم تشجيع البحث في فصل نظائر الليزر لتخصيب اليورانيوم (Kemp, 2012) . إن أية تقنية تجعل من الممكن إنتاج مواد انشطارية تُستخدم في صنع الأسلحة باستخدام طاقة أقل ، أو بصمة صناعية أصغر ، من شأنها أن تؤدي إلى تآكل الحواجز المهمة أمام الانتشار. فمن الصعب أن نرى كيف سيتم تعويض التخفيض الطفيف في أسعار الطاقة النووية. على العكس من ذلك ، ربما يكون العالم أفضل حالًا إذا أصبح تخصيب اليورانيوم بطريقة ما أكثر صعوبة وأكثر تكلفة. ما نريده ، بشكل مثالي ، في هذا المجال ، ليس التقدم التكنولوجي لكن الانحدار التكنولوجي.

وفي حين أن الانحدار المستهدف قد لا يكون في البرامج ، يمكننا أن نهدف إلى إبطاء معدل التقدم نحو التقنيات التي تزيد من المخاطر بالنسبة إلى معدل التقدم في تقنيات الحماية. هذه هي الفكرة التي عبر عنها مبدأ التطور التكنولوجي التفضيلي. يركز المبدأ في صيغته الأصلية على الخطر الوجودي. ولكن يمكننا تطبيقه على نطاق أوسع ليشمل أيضًا التقنيات ذات الإمكانات المدمرة "فقط" :

مبدأ التطور التكنولوجي التفضيلي يعيق تطوير التقنيات الخطرة والضارة ، خاصة تلك التي ترفع من مستوى الخطر الوجودي ؛ وتسرع تطوير التقنيات الربحية ، وخاصة تلك التي تقلل من المخاطر الوجودية التي تشكلها الطبيعة أو التقنيات الأخرى (Bostrom, 2002) .

يتوافق مبدأ التطور التكنولوجي التفضيلي مع الأشكال المعقولة للحتمية التكنولوجية. على سبيل المثال ، حتى لو تقرر أن جميع التقنيات التي يمكن تطويرها سوف يتم تطويرها ، فلا يزال هذا المبدأ أمراً مهماً عند تطويرها. ويمكن أن يُحدث ترتيب الوصول إليها فرقًا مهماً – من الناحية المثالية ، يجب أن تأتي تقنيات الحماية قبل التقنيات المدمرة التي تهدف إلى الحماية من خطرها ؛ أو ، إذا لم يكن ذلك ممكنا ، فمن المستحسن أن يتم تقليل الفجوة إلى الحد الأدنى بحيث يمكن للتدابير المضادة الأخرى (أو الحظ) أن تدفعنا إلى أن نوفر الحماية القوية. يؤثر توقيت الاختراع أيضًا على السياق الاجتماعي السياسي الذي تولد فيه التكنولوجيا. على سبيل

المثال ، إذا كنا نعتقد أن هناك اتجاهًا علمانياً نحو جعل الحضارة أكثر قدرة على التعامل مع الكرات السوداء ، فقد نرغب في تأخير أكثر للتطورات التكنولوجية الخطرة ، أو على الأقل الامتناع عن تسريعها. حتى لو افترضنا أن الدمار الحضاري أمر لا يمكن تجنبه ، فإن الكثيرين يفضلون أن يحدث في المستقبل ، في وقت ربما لم يعودوا هم وأحباؤهم أحياء على أية حال.(٣٢)

التطور التكنولوجي التفضيلي ليس له معنى حقًا في نموذج الجرة الإبداع الأصلي ، حيث يأتي لون كل كرة مفاجأة كلية. إذا أردنا استخدام نموذج الجرة في هذا السياق ، فيجب علينا تعديله. يمكننا أن نشترط ، على سبيل المثال ، أن يكون للكرات قوام مختلف وأن يوجد ارتباط بين الملمس واللون ، حتى تكون لدينا أدلة حول لون الكرة قبل أن نستخرجها. هناك طريقة أخرى لجعل الاستعارة أكثر واقعية وهي تخيل وجود خيوط أو أشرطة مرنة بين بعض الكرات ، بحيث أننا عندما نسحب إحداها نسحب بذات الفعل العديد من الكرات الأخرى التي ترتبط بها. من المفترض أن تكون الجرة أسطوانية بدرجة كبيرة ، حيث يجب أن تظهر تقنيات معينة قبل الوصول إلى تقنيات أخرى (من غير المحتمل أن نجد مجتمعًا يستخدم الطائرات النفاثة والفؤوس الحجرية). كما أن الاستعارة سوف تصبح أكثر واقعية إذا تخيلنا أنه لا توجد يد واحدة فقط تستكشف الجرة بلطف: بدلاً من ذلك ، تخيل حشدًا من الملثمين يتشاجرون بكل طاقاتهم على أمل الحصول على الذهب والمجد والاستشهاد.

من الواضح أن التنفيذ الصحيح للتطور التكنولوجي التفضيلي يعد مهمة استراتيجية صعبة، راجع (Collingridge, 1980) ومع ذلك ، بالنسبة للجهة الفاعلة التي تهتم بإيثار النتائج طويلة الأجل والتي تشارك في بعض المشاريع الإبداعية (على سبيل المثال كباحث ، أو ممول ، أو رائد أعمال ، أو منظم ، أو مشرِّع) فإن الأمر يستحق المحاولة. بعض التطبيقات ، على أي حال ، تبدو واضحة إلى حد ما: على سبيل المثال ، لا تعمل على فصل نظائر الليزر ، ولا تعمل على الأسلحة البيولوجية ، ولا تطور أشكالًا من الهندسة الجيولوجية التي من شأنها تمكين الأفراد العشوائيين من إجراء تعديلات جذرية من جانب واحد على مناخ الأرض، فكر مليًا قبل تسريع تقنيات التمكين – مثل آلات تخليق الحمض النووي – التي من شأنها أن تسهل بشكل مباشر مثل هذه التطورات المشؤومة. (٣٣) ولكن اعمل على تعزيز التقنيات التي يغلب

عليها الطابع الوقائي ؛ على سبيل المثال ، تلك التي تتيح رصدًا أكثر كفاءة لتفشي الأمراض أو تلك التي تسهل اكتشاف برامج أسلحة الدمار الشامل السرية.

حتى إذا كان من المحتم تطوير جميع التقنيات "السيئة" الممكنة في نهاية المطاف ، فلا يزال من المفيد شراء القليل من الوقت. (٣٤) ومع ذلك ، فإن التطور التكنولوجي التفضيلي لا يقدم بمفرده حلاً لنقاط الخطورة المحتملة التي لا تزال قائمة لفترات طويلة – تلك التي يصعب فيها تطوير تقنيات الحماية الكافية مقارنة بنظيراتها المدمرة ، أو التي يتمتع فيها التدمير بما يميزه حتى عند النضج التكنولوجي. (٣٥)

تعديل التفضيل **Preference modification**

هناك طريقة أخرى ممكنة نظريًا لتحقيق الاستقرار الحضاري ، وهي تغيير حقيقة وجود عدد كبير من الفاعلين الذين يمثلون توزيعا بشريا واسعا ومعروفا للدوافع. إننا نحتفظ للمناقشة لاحقا بالتدخلات التي من شأنها أن تقلل من العدد الفعال للجهات الفاعلة المستقلة من خلال زيادة أشكال التنسيق المختلفة. هنا ننظر في إمكانية تعديل توزيع التفضيلات (ضمن مجموعة ثابتة إلى حد ما من الفاعلين) . تعتمد الدرجة التي يبشر بها هذا النهج على نوع الخطورة المحتملة الذي نضعه في الاعتبار.

في حالة الخطورة المحتملة من النوع ١ ، لا يبدو تعديل التفضيل واعدًا ، على الأقل في غياب وسائل فعالة للغاية للقيام بذلك. ضع في اعتبارك أن بعض نقاط الخطورة المحتملة من النوع ١ قد تؤدي إلى دمار حضاري إذا كان هناك ثمة شخص واحد مفوَّض في أي مكان في العالم لديه الدافع للسير باتجاه النتيجة المدمرة. مع هذا النوع من الخطورة المحتملة ، لن يكون مفيدا تقليل عدد الأشخاص ، الذين سيشهدون الحادث المروع ، في منع الدمار ما لم يتم تقليل العدد إلى الصفر ، وهو ما قد يكون غير ممكن تمامًا. صحيح أن هناك خطورات أخرى محتملة من النوع ١ والتي قد تتطلب عددا أكبر من الضحايا ، إلى حد ما ، من أجل حدوث الدمار الحضاري: على سبيل المثال ، في سيناريو مثل "الأسلحة النووية السهلة" ، ربما يجب أن يكون هناك شخص ما من بقايا الحادث المروع في كل واحدة من عدة مئات من المدن. لكن هذا لا يزال مستوىً منخفضًا جدًا. من الصعب تخيل التدخل – دون إعادة هندسة الطبيعة البشرية

جذريًا على نطاق عالمي بالكامل – ذلك من شأنه أن يستنفد عدد ضحايا الحادث المروع بشكل كافٍ للقضاء تمامًا أو حتى تقليل خطر نقاط الخطورة المحتملة من النوع الأول.

لاحظ أن التدخل الذي يخفض حجم ضحايا الحادث المروع إلى النصف لن يقلل (على الأقل ليس من خلال أي تأثير أولي) من المخاطر المتوقعة من مصادر الخطورة المحتملة من النوع ١ في أي مكان قريب بنفس القدر. سيكون تخفيض خطر النوع الأول بنسبة ٥% أو ١٠% أكثر معقولية من خفض ضحايا الحادث المروع إلى النصف. والسبب هو أن هناك شكًا واسعًا حول مدى تدمير بعض تقنيات الكرات السوداء الجديدة ، ويمكن القول أننا يجب أن نستخدم وتيرة موحدة بشكل مسبق إلى حد ما في مساحة التسجيل (على مدى عدة أمثال من حيث الحجم) من حجم ضحايا الحادث المروع ، الذي سيكون مطلوبا حتى يحدث الدمار الحضاري ، بشرط ظهور خطورة محتملة من النوع الأول. بعبارة أخرى ، بشرط تطوير بعض التقنيات الجديدة التي تجعل من السهل على الفرد العادي قتل مليون شخص على الأقل ، وقد يكون من المحتمل (تقريبًا) أن تمكِّن التكنولوجيا الفرد العادي من قتل مليون شخص ، أو عشرة ملايين من البشر ، أو مائة مليون شخص ، أو مليار شخص ، أو كل إنسان على قيد الحياة.

على الرغم من هذه الاعتبارات ، يمكن أن يكون تعديل التفضيل مفيدًا ، في السيناريوهات التي تكون فيها مجموعة الفاعلين الفوَّضين محدودة في البداية ببعض المجموعات السكانية الفرعية الصغيرة ، التي يمكن تحديدها. قد يكون من الصعب استخدام بعض تقنيات الكرة السوداء ، عند ظهورها لأول مرة من الجرة ، وتتطلب معدات متخصصة. قد تكون هناك فترة من عدة سنوات قبل أن يتم إتقان مثل هذه التكنولوجيا لدرجة أن الفرد العادي يمكن أن يتقنها. خلال هذه الفترة المبكرة ، يمكن أن تكون مجموعة الجهات الفاعلة التي تم تمكينها محدودة إلى حدا ما؛ على سبيل المثال ، قد تتكون حصريًا من أفراد لديهم خبرة في العلوم الحيوية يعملون في نوع معين من المعامل. يمكن للفحص الدقيق للمتقدمين لشغل وظائف في مثل هذه المختبرات أن يُحدث تأثيرًا ذا مغزى في المخاطرة ، لدرجة أن يتمكن الفرد المدمر من الوصول إلى الكرة السوداء للتكنولوجيا الحيوية في غضون السنوات القليلة الأولى من ظهورها. (٣٦) وقد يوفر هذا التأجيل فرصة لتقديم تدابير مضادة أخرى لتوفير مزيد من الاستقرار الدائم ، تحسبًا للوقت الذي تصبح فيه التكنولوجيا سهلة بما يكفي لاستخدامها بحيث تنتشر بين عدد أكبر من السكان.

بالنسبة لأنواع الخطورة المحتملة من النوع ٢ أ ، فإن مجموعة الجهات الفاعلة المفوَّضة أصغر بكثير. ما نتعامل معه هنا عادة هو الدول ، ربما إلى جانب عدد قليل من الجهات الفاعلة القوية غير الحكومية. في بعض سيناريوهات الخطورة من النوع ٢ أ ، قد تتكون المجموعة حصريًا من قوتين عظميين ، أو حفنة من الدول ذات القدرات الخاصة (كما هو الحال حاليًا مع الأسلحة النووية) . وبالتالي ، قد يكون من المفيد جدًا تغيير تفضيلات حتى بعض الدول القوية لاتخاذ توجها أكثر سلامًا. سوف يكون سيناريو "الضربة الأولى الآمنة" أقل إثارة للقلق إذا كان لدى الجهات الفاعلة ، التي تواجه معضلة أمنية ، مواقف تجاه بعضها البعض مماثلة لتلك السائدة بين فنلندا والسويد. وبالنسبة للكثير من المجموعات المعقولة من الحوافز ، التي يمكن أن تظهر لدى الجهات الفاعلة القوية نتيجة لبعض الاختراقات التكنولوجية ، فإن الاحتمالات المؤدية إلى الحصول على نتيجة غير مدمرة يمكن تجميلها بشكل كبير ، إذا كان لدى الجهات الفاعلة المعنية نزعات أكثر عدوانية. على الرغم من أن هذا يبدو صعب التحقيق ، إلا أنه ليس من الصعب إقناع جميع الأعضاء تقريبًا من ضحايا الانفجار المروع بتغيير تصرفاتهم.

أخيرًا ، ضع في اعتبارك النوع ٢ ب من الخطورة المحتملة. تذكر أن مثل هذه الخطورة تستلزم "بشكل افتراضي" أن عددًا كبيرًا من الجهات الفاعلة يواجه دوافع اتخاذ بعض الإجراءات المدمرة ، بحيث تضيف التأثيرات المجمعة إلى الدمار الحضاري. لذلك يجب أن تكون الحوافز الدافعة لاستخدام تكنولوجيا الكرة السوداء هي تلك التي تسيطر على جزء كبير من سكان العالم – وربما تكون المكاسب الاقتصادية هي المثال الرئيسي لمثل هذا الدافع شبه العالمي. لذا تخيلوا إجراءً خاصاً من نوع معين ، متاحا لكل فرد تقريبًا ، والذي يوفر لكل شخص يتخذه جزءًا صغيرًا X من دخله أو دخلها السنوي ، بينما تَنتُج عوامل خارجية سلبية بحيث إذا قام نصف سكان العالم بهذا الإجراء فإن الحضارة تدمَّر. عندما إكس = صفراً يمكننا أن نفترض أن عددًا قليلاً من الأشخاص سوف يتخذون الإجراء المعادي للمجتمع. ولكن كلما كانت X أكبر زادت نسبة السكان التي من شأنها أن تستسلم للإغراء. لسوء الحظ ، من المعقول أن تكون قيمة X التي من شأنها أن تحث نصف السكان على الأقل على اتخاذ الإجراء نسبة صغيرة ، ربما أقل من ١% (٣٧) بينما سيكون مرغوبا أن نغير توزيع التفضيلات العالمية لكي نجعل الناس أكثر إيثارًا

ويرفعون قيمة X ، وهذا يبدو صعب التحقيق. (ضع في اعتبارك القوى القوية الكثيرة التي تتنافس بالفعل على القلوب والعقول – شركات الإعلان ، والمنظمات الدينية ، والحركات الاجتماعية ، وأنظمة التعليم ، وما إلى ذلك.) حتى الزيادة الهائلة في مقدار الإيثار في العالم – المقابِلة ، دعنا نقول ، لمضاعفة X من ١% إلى ٢% – ستمنع وقوع الكارثة فقط في نطاق ضيق نسبيًا من السيناريوهات ، بالتحديد تلك التي تكون فيها الفائدة الخاصة لاستخدام التكنولوجيا المدمرة تتراوح في نطاق ١ إلى ٢ في المائة. فالسيناريوهات التي تتجاوز فيها المكاسب الخاصة ٢ في المائة سوف تؤدي ، على أي حال ، إلى تدمير حضاري.

باختصار ، قد يكون تعديل توزيع التفضيلات خلال مجموعة الجهات الفاعلة ، التي سيتم تمكينها تدمريا من خلال اكتشاف الكرة السوداء ، بمثابة مساعد مفيد لوسائل الاستقرار الأخرى ، ولكن قد يكون من الصعب تنفيذه ولن يقدم ، في أفضل الأحوال ، سوى حماية جزئية للغاية (ما لم نفترض أشكالًا قصوى من إعادة هندسة الطبيعة البشرية في جميع أنحاء العالم.) . (٣٨)

بعض التدابير المضادة المحدَّدة وقيودها

Some specific countermeasures and their limations

إلى جانب التأثير على اتجاه التقدم العلمي والتكنولوجي ، أو تغيير التفضيلات المتعلقة بالتدمير ، هناك مجموعة متنوعة من الإجراءات المضادة الممكنة الأخرى ، التي يمكن أن تخفف من الخطورة الحضارية. على سبيل المثال ، يمكن للمرء أن يحاول:

– منع انتشار المعلومات الخطرة ؛

– تقييد الوصول إلى المواد المطلوية ، والأدوات ، والبنية التحتية ؛ و

– ردع الأشرار المحتملين عن طريق زيادة فرص القبض عليهم ؛ و

– أن نكون أكثر حذرا مع القيام بالمزيد من أعمال تقييم المخاطر ؛ و

– إنشاء نوع من آليات المراقبة والتنفيذ ، التي من شأنها أن تجعل من الممكن منع محاولات القيام بعمل هدَّام.

ينبغي أن يكون واضحًا من مناقشتنا والأمثلة السابقة أن الأربعة الأولى من هذه الحلول ليست حلولًا عامة. قد يكون من السهل منع انتشار المعلومات. حتى لو أمكن القيام بذلك ، فلن يمنع ذلك من إعادة اكتشاف المعلومات الخطرة بشكل مستقل. يبدو أن الرقابة في أفضل الأحوال تعد تدبيرا مؤقتا. (٣٩) ويُعد تقييد الوصول إلى المواد ، والأدوات ، والبنية التحتية طريقة رائعة للتخفيف من بعض أنواع التهديدات (الكرة الرمادية) ، ولكنها غير مفيدة لأنواع أخرى من التهديدات – مثل تلك التي تتطلب المكونات المطلوبة في أماكن كثيرة جدًا في الاقتصاد ، أو تكون متاحة بالفعل في كل مكان ، عند اكتشاف الفكرة الخطيرة (مثل الزجاج ، والمعدن ، والبطاريات في سيناريو "الأسلحة النووية السهلة"). إن ردع الأشرار المحتملين أمر منطقي ؛ ولكن بالنسبة للتقنيات المدمرة بما فيه الكفاية ، ووجود ضحايا الانفجار المروع يجعل الردع غير كافٍ ، حتى لو كان من المؤكد القبض على كل مرتكبي الجريمة.

إن ممارسة المزيد من الحذر والقيام بمزيد من تقييم المخاطر تعد أيضاً استراتيجية ضعيفة ومحدودة. قد لا يساعدنا كثيرا أن يكون أحد الفاعلين أكثر حذرا من جانب واحد في ما يتعلق بالخطورة المحتملة من النوع ٢ أ ، ولن يؤثر بشيء في الأساس بالنسبة لخطورة من النوع ٢ ب أو النوع ١ . في حالة الخطورة من النوع صفر ، قد يكون من المفيد أن يكون الفاعل المحوري أكثر حذرًا – هذا لو أن الفاعل الأول الذي يسير بخطى حذرة ، فقط ، لم يُتْبَعْ باندفاع من الجهات الفاعلة غير الحذرة للوصول إلى نفس التكنولوجيا الخطرة (ما لم يكن العالم بطريقة ما ، في غضون ذلك ، قد حقق الاستقرار بوسائل أخرى) . (٤٠) وفيما يتعلق بتقييم المخاطر ، يمكن أن يقلل الخطر ، فقط ، إذا أدى إلى تنفيذ بعض الإجراءات المضادة الأخرى. (٤١)

يشير الإجراء المضاد الأخير في القائمة – المراقبة – إلى حل أكثر عمومية. وسوف نناقشه في القسم التالي تحت عنوان "الشرطة الوقائية" . لكن يمكننا بالفعل أن نلاحظ أنها لا تكفي وحدها. على سبيل المثال ، بالنظر إلى الخطورة المحتملة من النوع ٢ ب مثل "الاحتباس الحراري العالمي الأسوأ" . حتى لو أتاحت المراقبة لدولة ما أن تنفذ بشكل كامل أي تنظيم بيئي تختار فرضه ، فلا تزال هناك مشكلة في جعل عدد كافٍ من الدول يوافق على اعتماد اللوائح المطلوبة – وهو أمر قد لا يحدث بسهولة. تتجلى حدود المراقبة بشكل أكثر وضوحًا في حالة الخطورة من النوع ٢ أ ، مثل "الضربة الأولى الآمنة" ، حيث تكمن المشكلة في أن الدول (أو

الجهات الفاعلة القوية الأخرى) يتم تحفيزها بشدة للقيام بأعمال مدمرة. إن قدرة تلك الدول على السيطرة الكاملة على ما يجري داخل حدودها لا تحل هذه المشكلة. الشيء المطلوب لحل المشكلات ، التي تنطوي على تحديات التنسيق الدولي بشكل موثوق ، هو الحوكمة العالمية الفعالة.

ثغرات الحوكمة **Governance gaps**

يجب أن تكون القيود المفروضة على التخلي التكنولوجي ، وتعديل التفضيل ، ومختلف التدابير المضادة المحددة واضحة الآن كردود على الخطورة الحضارية المحتملة. لذلك ، إلى الحد الذي نشعر فيه بالقلق من أن فرضية العالم المعرض للخطر VWH قد تكون صحيحة ، يجب علينا النظر في الطريقتين المحتملتين المتبقيتين لتحقيق الاستقرار:

١- خلق القدرة على جعل الشرطة الوقائية فعالة للغاية. تطوير قدرة الحوكمة اللازمة داخل الدولة لمنع أي فرد أو مجموعة صغيرة – بما في ذلك تلك التي لا يمكن ردعها – بمصداقية عالية للغاية – من تنفيذ أي إجراء غير قانوني بدرجة كبيرة ؛ و

٢- خلق القدرة على إقامة الحوكمة العالمية القوية. تطوير قدرة الحوكمة بين الدول اللازمة لحل المشكلات العالمية الشائعة الأكثر خطورة بشكل موثوق وضمان التعاون القوي بين الدول (والمنظمات القوية الأخرى) حيثما تكون المصالح الأمنية الحيوية على المحك – حتى عندما تكون هناك دوافع قوية للغاية لتشويه الاتفاقات ، أو رفض التوقيع في المقام الأول.

الفجوتان الحوكميتان المنعكستان في (١) و (٢) ، إحداهما على المستوى الجزئي ، والأخرى على النطاق الكلي ، هما كعبا (مثنى كعب) أخيل في النظام العالمي المعاصر. وطالما بقيا غير محميَّيْن ، فإن الحضارة تظل عرضة للكرة السوداء التكنولوجية المحتملة ، التي من شأنها أن تمكِّن من توجيه الضربة إليها. ما لم يظهر مثل هذا الاكتشاف من الجرة ، وإلى أن يتم ذلك ، فمن السهل التغاضي عن مدى كوننا عرضة لهذا الخطر.

في القسمين التاليين ، سنناقش كيف أن سد فجوات الحوكمة هذه ضروري لتحقيق قدرة عامة على استقرار نقاط الخطورة الحضارية المحتملة. وغني عن البيان أن هناك صعوبات كبيرة، وأيضًا جوانب سلبية محتملة خطيرة للغاية ، في السعي لتحقيق التقدم نحو (١) و (٢) .

في هذه الورقة ، سنقول القليل عن الصعوبات ولن نقول شيئًا تقريبًا عن الجوانب السلبية المحتملة – جزئيًا لأنها معروفة جيدًا بالفعل وتحظى بالتقدير على نطاق واسع. ومع ذلك ، فإننا نؤكد أن عدم وجود مناقشة حول الحجج ضد (١) و (٢) لا ينبغي تفسيره على أنه تأكيد ضمني على أن هذه الحجج ضعيفة ، أو أنها لا تشير إلى أمور ذات أهمية. بالطبع ، يجب أن تؤخذ في الاعتبار في تقييم شامل لكل الأشياء. لكن مثل هذا التقييم يقع خارج نطاق المساهمة الحالية ، والتي تركز بشكل خاص على الاعتبارات المنبعثة من فرضية العالم المعرض للخطر VWH.

الشرطة الوقائية **Preventing policing**

نفترض أن خطورة محتملة من النوع ١ انفتحت. فيكتشف شخص ما طريقة سهلة حقًا لإحداث دمار شامل. تنتشر المعلومات حول الاكتشاف. المواد والأدوات المطلوبة متوفرة في كل مكان ولا يمكن إزالتها بسرعة من نطاق التداول. بالطبع من غير القانوني تمامًا لأي جهة فاعلة غير حكومية تدمير مدينة ، وأي شخص يتم القبض عليه وهو يفعل ذلك سيتعرض لعقوبات قاسية. لكن من المعقول أن أكثر من شخص واحد من بين المليون ينتمي إلى ضحايا الانفجار المروع الذي لا يمكن ردعه. على الرغم من صغر حجمه نسبيًا ، إذا كان كل شخص من هذا القبيل يخلق حدثًا مدمرًا لمدينة ، فإن العدد المطلق لا يزال أكبر من أن تتحمله الحضارة. اذن ، ما العمل؟

إذا وجدنا أنفسنا فجأة في مثل هذا الموقف ، فقد يكون قد فات الأوان لمنع تدمير الحضارة. ومع ذلك ، من الممكن تصور سيناريوهات يستطيع فيها المجتمع البشري البقاء على قيد الحياة رغم هذا التحدي دون تغيير – والتحدي الأكثر صعوبة يكمن حيث يمكن للأفراد كل بمفرده تدمير ، ليس فقط مدينة واحدة ، ولكن ، العالم بأسره.

ما هو مطلوب لتحقيق الاستقرار في مثل هذه الخطورات المحتملة هو قدرة الشرطة الوقائية المتطورة للغاية. سوف تحتاج الدول إلى القدرة على مراقبة مواطنيها عن كثب بما يكفي للسماح لهم باعتراض أي شخص يبدأ في التحضير لعمل دمار شامل.

تعتمد جدوى مثل هذه المراقبة والاعتراض على تفاصيل السيناريو: كم يستغرق من الوقت لنشر تقنية الكرة السوداء بشكل مدمر؟ ما مدى إمكانية ملاحظة الإجراءات المعنية؟ هل يمكن

تمييزها من السلوك الذي لا نريد منعه؟ ولكن من المعقول أن يتم تثبيت جزء كبير من الخطورة المحتملة من النوع الأول من خلال حالة انتشار التقنيات المتاحة حاليًا إلى أقصى حد. وستؤدي التطورات المتوقعة في تكنولوجيا المراقبة إلى توسيع نطاق الحماية القابلة للتحقيق بشكل كبير.

لتكوين صورة لما يمكن أن يبدو عليه مستوى المراقبة المكثفة حقًا ، تأمل المقالة القصيرة التالية:

المراقبة الشاملة Panopticon عالية التقنية (****)

كل شخص متناسب مع "بطاقة الحرية" – وهي تتبع أجهزة المراقبة التي يمكن ارتداؤها والأقل شيوعا اليوم ، مثل بطاقة الكاحل المستخدمة في عدد من البلدان كبديل عن السجن ، وكاميرات الجسم التي يرتديها الكثير من قوات الشرطة ، وأجهزة تعقب الجيب وأساور المعصم التي يستخدمها بعض الآباء لتتبع أطفالهم ، وبالطبع ، الهاتف الخلوي في كل مكان (والذي تم وصفه بأنه "جهاز تتبع شخصي يمكن استخدامه أيضًا لإجراء المكالمات") . (٤٢) بطاقة الحرية هي جهاز أكثر تقدمًا بدرجة قليلة ، يتم ارتداؤها حول الرقبة ومزينة بكاميرات وميكروفونات متعددة الاتجاهات. يتم تحميل ملفات الفيديو والصوت المشفرة باستمرار من الجهاز إلى الجهاز الرئيسي وتتجري ترجمته آليا في الوقت الفعلي. تصنِّف خوارزميات الذكاء الاصطناعي أنشطة الشخص الذي يرتديها ، وحركات يده ، والأشياء القريبة ، وغيرها من الإشارات الظرفية. إذا تم الكشف عن نشاط مشبوه ، فالخلاصة تُنقل إلى واحدة من عدة محطات مراقبة باتريوت. وهي مجمعات مكاتب واسعة يجري العمل بها بصفة دائمة ٢٤ ساعة ٧ أيام في الأسبوع. هناك ، يراجع ضابط الحرية بث الفيديو على عدة شاشات ويستمع إلى الصوت في سماعات الرأس. ثم يحدد ضابط الحرية الإجراء المناسب ، مثل الاتصال بمرتدي العلامة tagwearer عبر رابط صوتي لطلب توضيحات أو لطلب رؤية أفضل. يمكن لضابط الحرية أيضًا إرسال مفتش أو وحدة استجابة سريعة من الشرطة أو طائرة بدون طيار drone لمزيد من التحقيق. في الجزء الصغير من الحالات التي يرفض فيها مرتديها التوقف عن النشاط المحظور بعد تحذيرات متكررة ، قد يتم إلقاء القبض عليه أو فرض عقوبات مناسبة أخرى. لا يُسمح للمواطنين بإزالة علامة الحرية ، إلا أثناء وجودهم فيها البيئات التي تم تجهيزها بأجهزة استشعار خارجية مناسبة (والتي تشمل مع ذلك معظم البيئات الداخلية والسيارات) . يوفر النظام حماية خصوصية متطورة

إلى حد ما ، مثل التشويش التلقائي لأجزاء الجسم الحميمية ، ويوفر خيارًا لتنقيح البيانات التي تكشف عن الهوية مثل الوجوه وبطاقات الأسماء وإصدارها فقط عند الحاجة إلى المعلومات لإجراء تحقيق. تراقب كل الآليات التي يدعمها الذكاء الاصطناعي والإشراف البشري عن كثب جميع الإجراءات التي يتخذها ضباط الحرية لمنع الانتهاكات. (٤٣)

سوف يتطلب إنشاء وتشغيل المراقبة الشاملة عالية التقنية استثمارات كبيرة ، ولكن بفضل انخفاض أسعار الكاميرات ، ونقل البيانات ، والتخزين ، والحوسبة ، والتقدم السريع في تحليل المحتوى المدعوم بالذكاء الاصطناعي ، فقد يصبح خلال فترة قصيرة ممكنًا تقنيًا وبأسعار معقولة. على سبيل المثال ، إذا انخفضت تكلفة تطبيق هذا على فرد واحد لمدة عام واحد إلى حوالي ١٤٠ دولارًا أمريكيًا ، فيمكن عندئذٍ مراقبة سكان العالم بالكامل بتكلفة أقل من ١ في المائة من الناتج المحلي الإجمالي العالمي. عند هذا السعر ، سيمثل النظام بشكل معقول مدخرات صافية – حتى مع تنحية استخدامه في منع الكوارث على نطاق الحضارة – بسبب فائدته في إنفاذ القانون بانتظام. إذا كان النظام يعمل كما هو معلن ، يمكن القضاء على الكثير من أشكال الجريمة تقريبًا ، مع ما يصاحب ذلك من تخفيضات في تكاليف الشرطة ، والمحاكم ، والسجون ، وأنظمة الأمن الأخرى. قد يؤدي أيضًا إلى نمو العديد من الممارسات الثقافية المفيدة التي يعوقها حاليًا نقص الثقة الاجتماعية.

إذا كانت الحواجز التقنية أمام المراقبة الشاملة عالية التقنية تتراجع بسرعة ، فماذا عن جدواها السياسية؟ أحد الاحتمالات هو أن المجتمع ينجرف تدريجياً نحو الشفافية الاجتماعية الكاملة حتى مع غياب أية صدمة كبيرة للنظام. قد يصبح من السهل بشكل تدريجي جمع وتحليل المعلومات حول الأشخاص والأشياء ، وقد يكون من الملائم جدًا السماح بذلك ، إلى الحد الذي يصبح فيه شيء قريب من المراقبة الكاملة حقيقة واقعة – وهذا قريب بما فيه الكفاية مع تحول آخر فقط يمكن تحويل المسار إلى مراقبة شاملة عالية التقنية. (٤٤) الاحتمال البديل هو أن نوعا معينا من الخطورة المحتملة الخاصة من النوع ١ تأتي بشكل صارخ بما يكفي لإخافة الدول ودفعها نحو اتخاذ تدابير قصوى ، مثل إطلاق برنامج صادم من أجل إنشاء مراقبة عالمية. قد تشمل التدابير القصوى الأخرى التي يمكن تجربتها في حالة عدم وجود نظام مراقبة عالمية شاملة تتضمن اعتماد سياسة الحبس الوقائي ، على سبيل المثال عندما تشير مجموعة من

المؤشرات غير الموثوقة إلى نسبة احتمالية تزيد عن ١ في المائة أن أحد الأفراد يحاول القيام بعمل من شأنه تدمير المدينة أو ما هو أسوأ. (٤٥) تعتمد المواقف السياسية تجاه مثل هذه السياسات على الكثير من العوامل ، بما في ذلك التقاليد الثقافية ومعايير الخصوصية والرقابة الاجتماعية ؛ لكنها سوف تعتمد أيضًا على مدى وضوح إدراك الخطورة الحضارية. على الأقل في حالة نقاط الخطورة المحتملة التي يوجد لها الكثير من الأحداث التحذيرية الواضحة ، فمن المعقول أن يتم إدراك الخطر بوضوح شديد. في سيناريو "الأسلحة النووية السهلة" ، على سبيل المثال ، بعد خراب عدد قليل من المدن الكبرى ، من المحتمل أن يكون هناك دعم شعبي قوي لسياسة قد تتضمن ، من أجل إحباط هجوم آخر ، حبس مائة من الأبرياء مقابل كل متآمر حقيقي. (٤٦) في مثل هذا السيناريو ، من المحتمل أن يتم دعم إنشاء مراقبة شاملة عالية التقنية على نطاق واسع باعتبارها أولوية ملحة للغاية. ومع ذلك ، بالنسبة لمواطن الخطورة المحتملة التي لم يسبقها أو يصاحبها مثل هذا الدليل الذي لا جدال فيه قد لا تتحقق أبدًا الرغبة في اتخاذ إجراءات وقائية قوية.

وبالتالي ، قد تكون الشرطة الوقائية الفعالة للغاية ، التي يتم تمكينها من خلال المراقبة في الوقت الحقيقي في كل مكان ، ضرورية لتحقيق الاستقرار في الخطورة المحتملة من النوع الأول. تعد المراقبة ذات صلة أيضًا ببعض أنواع الخطورة المحتملة الأخرى ، على الرغم من أنها ليست مركزية كما في حالة النوع ١ .

في الخطورة المحتملة من النوع ٢ ب ، تأتي النتيجة السيئة من خلال الإجراءات المشتركة لكتلة من الجهات الفاعلة المستقلة التي تم تحفيزها للتصرف بشكل مدمر. ولكن ما لم يكن من الصعب للغاية ملاحظة السلوكيات المدمرة ، فلن تكون هناك حاجة إلى تكثيف المراقبة أو الشرطة الوقائية لتحقيق الاستقرار. في حالة "الاحتباس الحراري العالمي الأسوأ" ، على سبيل المثال ، ليس من الضروري استباق الإجراءات الفردية. فالمستويات الخطرة للانبعاثات تستغرق وقتًا لكي تتراكم ، ويمكن محاسبة الملوثين بعد وقوع التلوث ؛ ويمكن تحمله إذا تسلل عدد قليل منهم عبر الاحتيال.

ومع ذلك ، بالنسبة لنقاط الخطورة المحتملة الأخرى من النوع ٢ ب ، ربما تكون أساليب المراقبة والرقابة الاجتماعية الموسَّعة مهمة. فكروا في "الغوغاء الجامحين" ، وهو سيناريو تتشكل

فيه مجموعة من الغوغاء التي تقتل أي شخص تتعامل معه ويرفض الانضمام إليهم ، والتي تنمو بشكل أكبر وأكثر قوة ، راجع (Munz et al., 2009) . إن السهولة التي يمكن أن تتشكل بها مثل هذه الموازين الاجتماعية السيئة وتنتشر ، وإمكانية إصلاحها بمجرد ترسخها ، والحصيلة التي تفرضها على رفاهية الإنسان ، تعتمد على معايير يمكن تغييرها من خلال الابتكارات التكنولوجية ، وربما يكون التغيير إلى الأسوأ. حتى في الوقت الحالي ، تكافح العديد من الدول لإخضاع الجريمة المنظمة. قد يؤدي اختراع الكرة السوداء (ربما نوع من التصميم الذكي لآلية التشفير الاقتصادي) الذي يجعل المؤسسات الإجرامية أكثر قابلية للتوسع أو أكثر ضررًا من حيث آثارها الاجتماعية ، إلى خلق خطورة محتملة لا يمكن تحقيق استقرارها إلا إذا امتلكت الدول قوى تكنولوجية غير مسبوقة للمراقبة والسيطرة الاجتماعية.

فيما يتعلق بأنواع الخطورة المحتملة من النوع ٢ أ ، حيث تنشأ المشكلة من الدوافع التي تواجه سلطات الدولة أو الجهات الفاعلة القوية الأخرى ، فليس من الواضح كيف يمكن أن تفيدنا المراقبة المألوفة. تاريخيًا ، ربما تكون الوسائل الأقوى للرقابة الاجتماعية قد أدت إلى تفاقم الصراع بين الدول – فقد اعتمدت الصراعات الأكثر دموية بين الدول على قدرات الحوكمة الفعالة للغاية للدولة الحديثة ، لتحصيل الضرائب ، والتجنيد الإجباري ، والدعاية للحرب. إنه لأمر وارد كون تحسين المراقبة يمكن أن يسهل بشكل غير مباشر في استقرار الخطورة المحتملة من النوع ٢ أ ، مثل تغيير الديناميكيات الاجتماعية الثقافية أو إنشاء خيارات جديدة لجعل معاهدات الحد من الأسلحة أو معاهدات عدم الاعتداء أكثر قابلية للتحقق. ولكن يبدو من المعقول بنفس القدر أن التأثير الواضح لتعزيز المراقبة المحلية وسلطات الشرطة على الخطورات المحتملة من النوع ٢ أ سيكون في الاتجاه المعاكس في حالة عدم وجود آليات موثوقة لحل النزاعات الدولية (أي يميل إلى إنتاج أو تفاقم الخطورات المحتملة هذه بدلاً من تحقيق الاستقرار فيها).

الحوكمة العالمية Global governance

عند التفكير مرة أخرى في "الضربة الأولى الآمنة": تواجه الدول التي تتمتع بإمكانية الوصول إلى تكنولوجيا الكرة السوداء بشكل افتراضي حوافز قوية لاستخدامها بشكل مدمر ، على الرغم من أنه سيكون من الأفضل للجميع ألا تفعل ذلك أية دولة. لقد تضمن المثال الأصلي

واقعًا مضادًا للأسلحة النووية ، ولكن بالنظر إلى المستقبل قد نحصل على هذا النوع من الكرة السوداء من التقدم في الأسلحة البيولوجية ، أو التصنيع الدقيق ذريًا ، أو إنشاء أسراب ضخمة من الطائرات القاتلة بدون طيار ، أو الذكاء الاصطناعي ، أو شيء آخر من هذا القبيل. ثم تواجه مجموعة الجهات الحكومية مشكلة العمل الجماعي. إن عدم حل هذه المشكلة يعني أن الحضارة تتعرض للدمار في هرمجدون نووية أو كارثة أخرى مماثلة. من المعقول أنه في غياب الحوكمة العالمية الفعالة ، سوف تفشل الدول في الواقع في حل هذه المشكلة. من خلال الافتراض ، فإن المشكلة التي تواجهنا هنا تطرح تحديات خاصة ؛ ومع ذلك ، فشلت الدول في كثير من الأحيان في حل مشكلات العمل الجماعي الأسهل. التاريخ البشري تغمره الآثار المرَضية للحرب من رأسه حتي إخمص قدميه.

ومع ذلك ، مع وجود إدارة عالمية فعالة ، يصبح الحل زهيداً: ببساطة منع جميع الدول من استخدام تقنية الكرة السوداء بشكل مدمر. في حالة "الضربة الأولى الآمنة" ، فإن الطريقة الأكثر وضوحًا للقيام بذلك هي الأمر بتفكيك جميع الأسلحة النووية وإنشاء نظام تفتيش ، مع أي مستوى من التدخلال ضروري لضمان عدم قيام أي شخص بإعادة إنشاء قدرة نووية. بدلا من ذلك ، مؤسسة الحوكمة العالمية يمكنها أن تحتفظ لنفسها بترسانة من الأسلحة النووية كحاجز ضد أي محاولة اختراق.

للتعامل مع الخطورة المحتملة من النوع ٢ أ ، ما تتطلبه الحضارة هو قدرة قوية على تحقيق التنسيق العالمي ، وتحديداً في الأمور التي تنطوي فيها تصرفات الدولة على عوامل خارجية كبيرة للغاية. من شأن الحوكمة العالمية الفعالة أن تساعد أيضًا في سيناريوهات النوع الأول والنوع ٢ ب حيث تُحْجِم بعض الدول عن إنشاء هذا النوع من الشرطة الوقائية التي قد تكون ضرورية لمنع الأفراد داخل أراضيها من القيام بعمل مدمر.

ضع في الاعتبار الكرة السوداء للتكنولوجيا الحيوية القوية بدرجة كافية بحيث يمكن أن يتسبب استخدام ضار واحد في حدوث جائحة من شأنها أن تقتل مليارات الأشخاص ، وبالتالي تمثل خطورة محتملة من النوع الأول. سوف يكون من غير المقبول حتى لو فشلت دولة واحدة في وضع الآلية اللازمة للمراقبة المستمرة والسيطرة على مواطنيها (أو أي آليات أخرى ضرورية لمنع الاستخدام الضار مع موثوقية كاملة بشكل افتراضي) . إن الدولة التي ترفض تنفيذ

الضمانات المطلوبة – ربما على أساس أنها تقدِّر الحرية الشخصية بشكل كبير أو تمنح المواطنين حق الخصوصية المنصوص عليه دستوريًا – سوف تكون عضوًا جانحًا في المجتمع الدولي. مثل هذه الدولة ، حتى لو كانت مؤسساتها الإدارية تعمل بشكل مثير للإعجاب في نواحٍ أخرى ، سوف تكون مماثلة لـ "الدولة الفاشلة" ، ومن ثم يجعلها افتقارها للسيطرة الداخلية ملاذاً آمناً للقراصنة والإرهابيين الدوليين (ومع ذلك فالمخاطر الخارجية التي تفرضها في هذه الحالة على بقية العالم ، بالطبع ، سوف تكون أكبر بكثير) . ومن المؤكد أن الدول الأخرى لديها أسباب للشكوى.

تنطبق حجة مماثلة على أنواع الخطورة المحتملة من النوع ٢ ب ، مثل سيناريو "الاحتباس الحراري العالمي الأسوأ" الذي تميل فيه بعض الدول إلى تجنب الجهود المكلفة التي يبذلها الآخرون لخفض الانبعاثات. يمكن لمؤسسة حوكمة عالمية فعالة أن تجبر كل دولة على القيام بدورها.

وهكذا نرى أنه في حين يمكن تثبيت بعض الخطورات المحتملة من خلال العمل الشرطي الوقائي وحده ، ويمكن تثبيت بعض نقاط الخطورة الأخرى من خلال الحوكمة العالمية وحدها ، إلا أن هناك بعض نقاط الخطورة التي تتطلب كليهما. سوف تكون الشرطة الوقائية شديدة الفعالية مطلوبة لأن الأفراد يمكن أن ينخرطوا في أنشطة يصعب تنظيمها والتي يجب مع ذلك تنظيمها بشكل فعال وقوي عالميًا ، أيضا سوف تكون الحوكمة العالمية القوية مطلوبة لأن الدول قد يكون لديها دوافع لعدم تنظيم هذه الأنشطة بشكل فعال حتى لو كانت لديها القدرة على القيام بذلك. ومع ذلك ، فإن عمل الشرطة الوقائية في كل مكان مدعوما من المراقبة والحوكمة العالمية الفعالة سوف يكونان كافييْن معًا لتحقيق الاستقرار في معظم نقاط الخطورة ، مما يجعل مواصلة التطوير العلمي والتكنولوجي آمنًا ، حتى لو كانت فرضية العالم المعرض للخطر VWH صحيحة. (٤٧)

وبالتالي ، فإن المراقبة الشاملة والحوكمة العالمية سوف توفران الحماية ضد مجموعة واسعة من نقاط الخطورة الحضارية. هذا سبب وجيه يؤيد تحقيق تلك الظروف المطلوبة. إن قوة هذا السبب تتناسب تقريبًا مع احتمالية صحة فرضية العالم المعرض للخطر.

وغني عن البيان أن الآلية التي تتيح أشكالًا مكثفة غير مسبوقة من المراقبة ، أو مؤسسة حوكمة عالمية ، قادرة على فرض إرادتها على أية دولة ، يمكن أن يكون لها أيضًا عواقب وخيمة؛ فالقدرات المحسنة للرقابة الاجتماعية يمكن أن تساعد الأنظمة الاستبدادية على حماية نفسها من التمرد. كما يمكن للمراقبة الشاملة أن تمكن أيديولوجية مهيمنة أو وجهة نظر أغلبية غير متسامحة من فرض نفسها على جميع جوانب الحياة ، مما يمنع الأفراد ذوي أنماط الحياة المنحرفة أو المعتقدات غير الشائعة من العثور على ملجأ يوفر لهم الاستقلال. وإذا اعتقد الناس أن كل ما يقولونه ويفعلونه هو ، بشكل فعال ، "مسجَّل" ، فقد يصبحون أكثر حذرًا ومتملقين بشكل مبتذل ، ويلتزمون بشكل وثيق بالنص القياسي للمواقف والسلوكيات الصحيحة سياسياً ، بدلاً من الجرأة على قول أو فعل أي شيء مثير من شأنه أن يجازف بجعلهم هدفًا للغوغاء الغاضبين ، أو وصمهم بعدم الأهلية الذي لا يُمحى مدى الحياة. الحوكمة العالمية ، من جانبها، يمكنها أن تقلل من الأشكال الربحية للمنافسة والتنوع بين الدول بما يخلق نظاما عالميا من نقطة واحدة من الفشل: إذا تم الاستيلاء على الحكومة العالمية من قبل أيديولوجية خبيثة أو مجموعة مصالح خاصة ، فقد تنتهي اللعبة من أجل التقدم السياسي ، لأن النظام الحالي قد لا يسمح أبدًا بإجراء تجارب باستخدام البدائل ، التي يمكن أن تكشف أن هناك طريقة أفضل. أيضًا ، إبتعادها عن متناول الأفراد والشعوب المتماسكة ثقافيًا أكثر مما عليه الآن حكومات الدول النموذجية ، قد يُنظر إلى مثل هذه المؤسسة من قبل البعض على أنها أقل شرعية ، وقد تكون أكثر عرضة لمشكلات ذاتية مثل التصلب البيروقراطي أو الانجراف السياسي بعيدًا عن المصلحة العامة.(٤٨)

وغني عن البيان أيضًا أن المراقبة القوية والحوكمة العالمية يمكن أن يكون لها عواقب جيدة مختلفة بصرف النظر عن استقرار نقاط الخطورة الحضارية ، انظر أيضا (Re, 2016; Bostrom, 2006; cf. Torres, 2018) . إن الأساليب الأكثر فعالية في الرقابة

الاجتماعية يمكن أن تقلل من الجريمة وتقلل من الحاجة إلى عقوبات جنائية قاسية. ويمكنها تعزيز مناخ من الثقة يمكِّن من ازدهار أشكال جديدة مفيدة من التفاعل الاجتماعي والنشاط الاقتصادي. ويمكن للحوكمة العالمية أن تمنع الحروب بين الدول ، بما في ذلك الحروب التي لا تهدد بتدمير الحضارة ، وتقلل من النفقات العسكرية ، وتعزز التجارة ، وتحل مختلف المشكلات البيئية العالمية وغيرها من المشكلات الشائعة ، وتهدئ من مشاعر القومية والكراهية والمخاوف ، وربما مع مرور الوقت قد تعزز الشعور الموسع بالتضامن العالمي. وقد ينتج عنها أيضًا زيادة التحولات الاجتماعية لفقراء العالم ، والتي قد يعتبرها البعض هدفا مرغوبا.

من الواضح أن هناك حجج قوية تؤيد أو تعارض التحرك في هذه الاتجاهات. وهذه الورقة لا تقدم أي حكم حول التوازن العام لهذه الحجج. فالطموح هنا أكثر محدودية: وهو توفير إطار عمل للتفكير في نقاط الخطورة الحضارية المحتملة المدفوعة بالتكنولوجيا ، والإشارة إلى أن القدرات الموسعة بشكل كبير للشرطة الوقائية والحوكمة العالمية سوف تكون ضرورية لتحقيق الاستقرار الحضاري في مجموعة من السيناريوهات. نعم ، يقدم التحليل سببًا إضافيًا لصالح تطوير هذه القدرات ، وهو سبب لا يبدو أنه كان يلعب دورًا مهمًا في العديد من المحادثات الأخيرة حول القضايا ذات الصلة ، مثل المناقشات حول الرقابة الحكومية وحول الإصلاحات المقترحة للمؤسسات الدولية والعابرة للقوميات. (٤٩) عند إضافة هذا السبب إلى الأسباب السابقة، فالتقييم يجب أن يصبح ، بالتالي ، أكثر ملاءمة مما كان محتملا أن يكون تجاه السياسات التي من شأنها أن تعزز قدرات الحكم بهذه الطرق. ومع ذلك ، فإن مسألة ما إذا كان هذا السبب الإضافي ذا قيمة بدرجة كافية لاستعادة التوازن العام أم لا فإن هذا يعتمد على اعتبارات أخرى تقع خارج نطاق هذه الورقة.

يجدر ، هنا ، التأكيد على أن الحجة الواردة في هذه الورقة تفضل أشكالا محددة لتقوية القدرة على الحكم. وفيما يتعلق بالمراقبة والشرطة الوقائية ، تشير اهتمامات فرضية العالم المعرض للخطر VWH ، على وجه التحديد ، إلى الرغبة في قدرة الحوكمة التي تجعل من الممكن بشكل موثوق للغاية قمع الأنشطة ، التي يرفضها بشدة عدد كبير جدًا من سكان هذا العالم (وأصحاب المصلحة المحليين المنوطين بالسلطة) . إنها توفر الدعم لأشكال أخرى من تعزيز الحوكمة ، فقط ، بقدر ما تساعد في خلق هذه القدرة الخاصة. وبالمثل ، فيما يتعلق

بالحوكمة العالمية ، تدعم الحجج المستندة إلى فرضية العالم المعرض للخطر VWH المؤسسات النامية القادرة على حل مشكلات التنسيق الدولي عالية المخاطر بشكل موثوق ، تلك التي يؤدي فيها الفشل في الوصول إلى حل إلى التدمير الحضاري. كما سيشمل ذلك امتلاك القدرة على منع صراعات القوى العظمى ، وقمع سباقات التسلح في أسلحة الدمار الشامل ، وتنظيم سباقات التطوير ونشر تقنيات الكرة السوداء المحتملة ، والقدرة على إدارة أسوأ أنواع المشكلات المأساوية الشائعة بنجاح. فهي لا تحتاج إلى القدرة على جعل الدول تتعاون في مجموعة من القضايا الأخرى ، ولا تشمل بالضرورة القدرة على تحقيق الاستقرار المطلوب باستخدام الوسائل المشروعة بالكامل فقط. في حين أن تلك القدرات قد تكون جذابة لأسباب أخرى ، فهي لا تظهر فورًا باعتبارها رغبات لمجرد أخذ فرضية العالم المعرض للخطر على محمل الجد. على سبيل المثال، فيما يتعلق بفرضية العالم المعرض للخطر ، سيكون من المُرضي نظريًا إذا تحققت قدرة الحوكمة العالمية المطلوبة من خلال صعود قوة عظمى واحدة إلى موقع الهيمنة الكافية لمنحها القدرة ، في حالة الطوارئ القصوى بما فيه الكفاية ، من جانب واحد على فرض خطة استقرار على بقية العالم. (*****)

إحدى القضايا المهمة التي ما زلنا بحاجة إلى مناقشتها هي مسألة التوقيت. حتى لو أصبحنا مهتمين للغاية بأن جرة الاختراع قد تحتوي على كرة سوداء ، فلا داعي لأن يدفعنا ذلك إلى تفضيل إنشاء مراقبة أقوى أو حوكمة عالمية الآن ، إذا اعتقدنا أنه سيكون من الممكن اتخاذ تلك الخطوات لاحقًا ، إذا ومتى ظهرت نقطة خطورة أمنية محتملة بوضوح. يمكننا عندئذٍ أن ندع العالم يواصل سباته الجميل ، في توقعات واثقة أنه بمجرد انطلاق المنبه ، سوف يقفز من الفِراش ويتخذ الإجراءات المطلوبة. لكننا يجب أن نتساءل عن مدى واقعية هذه الخطة.

من المفيد ، هنا ، استخدام بعض التأمل التاريخي. طوال الحرب الباردة ، عاشت القوتان العظميان (وكامل نصف الكرة الشمالي) في خوف مستمر من الإبادة النووية ، والتي كان من الممكن أن تنطلق في أي وقت عن طريق المصادفة ، أو نتيجة لبعض الأزمات التي خرجت عن نطاق السيطرة. لقد تم قبول حقيقة التهديد من قبل جميع الأطراف. كان من الممكن تقليل هذا الخطر بشكل كبير ، ببساطة ، عن طريق التخلص من جميع الأسلحة النووية أو معظمها (وهي خطوة كان من الممكن أيضا ، أن توفر أكثر من عشرة تريليونات دولار ، كتأثير جانبي

طفيف). (٥٠) (٥١) ومع ذلك ، وبعد عدة عقود من الجهود ، تم تنفيذ تدابير محدودة لنزع السلاح النووي وتدابير أخرى للحد من المخاطر. في الواقع ، لا يزال خطر الإبادة النووية قائما بيننا حتى يومنا هذا. في ظل غياب الحوكمة العالمية القوية التي يمكنها فرض معاهدة وإرغام المتنازعين على قبول حل وسط ، لم يتمكن العالم حتى الآن من حل مشكلة العمل الجماعي الأكثر وضوحًا. (٥٢)

لكن ربما يكون سبب فشل العالم ، في القضاء على خطر الحرب النووية ، هو أن الخطر لم يكن كبيرًا بما يكفي؟ لو كانت المخاطر أعلى ، لكان للمرء أن يجادل بدرجة من الابتهاج ، ولكان بالإمكان العثور على الإرادة اللازمة لحل مشكلة الحوكمة العالمية. ربما – على الرغم من أن هذا لا يبدو أساسا مستقرا إلى حد ما ليعتمد عليه مصير الحضارة. يجب أن نلاحظ أنه على الرغم من أن التكنولوجيا الأكثر خطورة من الأسلحة النووية قد تحفز إرادة أكبر للتغلب على العقبات التي تحول دون تحقيق الاستقرار ، فإن الخصائص الأخرى للكرة السوداء يمكن أن تجعل مشكلة الحوكمة العالمية أكثر صعوبة مما كانت عليه خلال الحرب الباردة. لقد أوضحنا بالفعل هذا الاحتمال في سيناريوهات مثل "الضربة الأولى الآمنة" و "الاحتباس الحراري العالمي الأسوأ". لقد رأينا كيف يمكن لخصائص معينة لمجموعة تقنية أن تولد حوافز أقوى للاستخدام المدمر أو لرفض الانضمام إلى (أو الانشقاق عن) أي اتفاق من أجل كبح تطبيقاتها الضارة.(٥٣)

حتى لو شعر المرء بالتفاؤل بإمكانية التوصل إلى اتفاق في نهاية المطاف ، فإن مسألة التوقيت يجب أن تظل مصدر اهتمام جاد. ويمكن لمشكلات العمل الجماعي الدولي ، حتى داخل مجال محدد ، أن تقاوم الحل طويل الأجل ، حتى عندما تكون المخاطر كبيرة ولا جدال فيها. يستغرق الأمر وقتًا لشرح سبب الحاجة إلى ترتيب والرد على الاعتراضات ، ووقتًا للتفاوض بشأن إنشاء تمثيل مقبول لدى الطرفين للفكرة التعاونية ، ووقت لوضع التفاصيل ، ووقتًا لإعداد الآليات المؤسسية المطلوبة للتنفيذ. في كثير من الحالات ، يمكن لمشكلات التعطيل والمعارضة الداخلية تأخير التقدم لعقود ؛ وفي الوقت الذي تكون فيه دولة متمردة جاهزة للانضمام ، فإن دولة أخرى كانت قد وافقت مسبقًا ربما تكون قد غيرت رأيها. ومع ذلك ، في الوقت نفسه ، يمكن أن تكون الفترة الفاصلة قصيرة بين ظهور إحدى الخطورات بشكل واضح للجميع والنقطة التي

يجب أن تكون عندها تدابير الاستقرار قائمة. بل يمكن أن تكون سلبية ، إذا تَركت طبيعة الخطورة مجالًا للإنكار أو إذا تعذر تقديم تفسيرات محددة على نطاق واسع بسبب مخاطر المعلومات. تشير هذه الاعتبارات إلى أنه من الصعب الاعتماد على تعاون دولي مخصص ، بشكل تلقائي ، لإنقاذ الموقف بمجرد ظهور خطورة محتملة. (٥٤)

يتشابه الوضع فيما يتعلق بالشرطة الوقائية من بعض النواحي ، على الرغم من أننا نرى اتجاهًا أسرع وأكثر قوة – مدفوعًا بالتقدم في تكنولوجيا المراقبة – نحو زيادة قدرات الدول على مراقبة تصرفات مواطنيها والسيطرة عليها أكثر من أي اتجاه نحو حوكمة عالمية فعالة. هذا صحيح على الأقل إذا نظرنا إلى العالم المادي. وفي عالم المعلومات الرقمية ، تبدو التوقعات أقل وضوحًا إلى حد ما ، نظرًا لانتشار أدوات التشفير وإخفاء الهوية ، وتكرار الابتكارات التخريبية التي تجعل من الصعب التنبؤ بمستقبل الفضاء الإلكتروني. ومع ذلك ، فإن القدرات القوية بما فيه الكفاية في الفضاء المادي سوف تثمر قدرات قوية في المجال الرقمي أيضًا. وفي أجهزة المراقبة عالية التقنية ، لن تكون هناك حاجة للسلطات لحل الشفرات ، حيث يمكنهم مراقبة كل شيء يكتبه المستخدمون في أجهزة الكمبيوتر الخاصة بهم وكل ما يظهر على شاشاتهم مباشرة.

يمكن للمرء أن يتخذ موقفًا مفاده أنه لا ينبغي تطوير الأساليب المحسّنة للمراقبة والرقابة الاجتماعية ما لم وإلى أن تظهر خطورة حضارية محددة بوضوح – وهو موقف يبدو خطيرا بما يكفي لتبرير التضحية ببعض أنواع الخصوصية وخطر التسهيل الطائش لكابوس شمولي. ولكن كما في حالة التعاون الدولي ، فإننا نواجه مسألة التوقيت. نظام مراقبة واستجابة متطور للغاية ، مثل تلك التي تم تصويرها في المراقبة عالية التقنية ، لا يمكن الحصول عليها وجعلها موثوقة تمامًا بين عشية وضحاها. من الناحية الواقعية ، من نقطة البداية الحالية ، سوف يستغرق تطبيق مثل هذا النظام سنوات عديدة ، ناهيك عن الوقت اللازم لبناء الدعم السياسي. إلا أن نقاط الخطورة المحتملة ، التي قد نحتاج لمثل هذا النظام للعمل ضدها ، ربما لا تقدم لنا الكثير من التحذير المسبق. في الأسبوع الماضي ، ربما نشر أحد أكبر المعامل الأكاديمية مقالًا في مجلة العلم Science وأثناء قراءتك لهذه الكلمات ، قد يقوم أحد المدونين المشهورين في مكان ما في العالم ، في سعي حثيث لمرات مشاهدة الصفحة ، بتحميل مشاركة تشرح طريقة ذكية

يمكن من خلالها لأي شخص استخدام نتيجة المختبر لإحداث دمار شامل. في مثل هذا السيناريو ، قد تحتاج الرقابة الاجتماعية المكثفة لأن يجري تفعيلها على الفور تقريبا. في سيناريو غير مواتٍ ، قد تكون المهلة قصيرة في حدود ساعات أو أيام. عندئذ يكون قد فات الأوان للبدء في تطوير بنية المراقبة عندما تظهر الخطورة المحتملة بوضوح. إذا أردنا تفادي الخراب ، فسوف يلزم وضع آلية الاستقرار بشكل مسبق.

ما قد يكون ممكنًا من الناحية النظرية هو تطوير القدرات للمراقبة المتطفلة والاعتراض في الوقت الحقيقي مسبقًا ، ولكن ليس لاستخدام هذه القدرات مبدئيا لتحقيق أي شيء مثل مداها الكامل. قد تكون هذه إحدى الطرق لتلبية متطلبات تحقيق الاستقرار في حالة الخطورة المحتملة من النوع الأول (ونقاط الخطورة المحتملة الأخرى التي تتطلب مراقبة موثوقة للغاية للإجراءات الفردية) . ومن خلال منح الحضارة البشرية القدرة على تكوين الشرطة الوقائية الفعالة للدرجة القصوى ، سوف نكون قد خرجنا من أحد الأبعاد شبه الفوضوية للحالة الافتراضية.

من المسلم به أن بناء مثل هذا النظام وإبقائه في وضع الاستعداد سوف يعني أن بعض الجوانب السلبية لإنشاء أشكال مكثفة للرقابة الاجتماعية سوف يتم تكبدها. على وجه الخصوص، هذا قد يجعل النتائج القمعية أكثر احتمالية:

"المسألة هي ما إذا كان إنشاء نظام للمراقبة يغير بشكل خطير هذا التوازن إلى حد كبير في اتجاه سيطرة الحكومة ... قد نتخيل نظامًا من الكاميرات الإجبارية المثبتة في المنازل ، يتم تفعيلها فقط بموجب أمر قضائي ، ويتم استخدامها مع الاحترام الصارم للقانون على مدى سنوات عديدة. تكمن المشكلة في أن مثل هذه البنية للمراقبة ، بمجرد إنشائها ، سيكون من الصعب تفكيكها ، وسوف تثبت أنها أداة تحكم قوية للغاية إذا وقعت في أيدي الأشخاص الذين – سواء من خلال الذعر ، أو الحقد ، أو الثقة المضللة في قدرتهم على الحكم سرًا على الصالح العام – سوف يسعون إلى استخدامها ضدنا." (Sanchez, 2013) .

إن تطوير نظام للشمولية الاستبدادية يعني مخاطرة لا علاج لها ، حتى لو لم يكن المرء ينوي تفعيلها.

يمكن للمرء أن يحاول تقليل هذه المخاطر من خلال تصميم النظام مزوَّداً بالضمانات الفنية والمؤسسية المناسبة. على سبيل المثال ، يمكن للمرء أن يهدف إلى نظام "شفافية منظَّمة" يمنع تركيز القوة من خلال تنظيم بنية المعلومات ، بحيث يجب على أصحاب المصلحة المستقلين المتعددين منح الإذن من أجل تفعيل النظام ، ولكي يجري توفير المعلومات المحددة فقط المطلوبة بشكل قانوني لأحد صانعي القرار ، وتطبيق إخفاء الهوية بما يسمح الغرض. مع تصميم آلية إبداعية من نوع محدد ، والتعلم الآلي بدرجة معينة ، ونوع من قواعد التشفير القوية، قد لا يكون هناك عائق أساسي أمام تحقيق نظام مراقبة فعال للغاية وعلى الفور في وظيفته الرسمية ، ولكنه أيضًا مقاوِم إلى حد ما للتخريب بالنسبة للاستخدامات البديلة.

ما مدى احتمالية تحقيق ذلك عمليًا ، هذه بالطبع مسألة أخرى ، تتطلب مزيدًا من الاستكشاف. (٥٥) حتى لو كان هناك خطر كبير من الشمولية سوف يصاحب حتما مشروع المراقبة حسن النية ، فلن يتبع ذلك متابعة مثل هذا المشروع الذي من شأنه أن يزيد من خطر الشمولية. إن المشروع حسن النية والأقل خطورة نسبيًا ، والذي بدأ في وقت يسوده الهدوء النسبي، قد يقلل من مخاطر الشمولية من خلال استباق مشروع أقل في حسن النية وأكثر خطورة والذي بدأ أثناء الأزمة. ولكن حتى لو كان هناك تأثير من نوع محدد خالٍ من زيادة مخاطر الشمولية ، فقد يكون من المفيد قبول هذا الخطر من أجل اكتساب القدرة العامة على تحقيق استقرار الحضارة في مواجهة التهديدات الناشئة من النوع الأول (أو من أجل الفوائد الأخرى التي يمكن أن تحققها الرقابة الفعالة للغاية والشرطة الوقائية) .

الاستنتاجات Conclusions

قدمت هذه الورقة منظورًا يمكننا من خلاله أن نرى بسهولة أكثر كيف تكون الحضارة عرضة لخطورة أنواع معينة من النتائج المحتملة لإبداعنا التكنولوجي – كما رسمنا كرة سوداء مجازية تخرج من جرة الاختراعات ، والتي لدينا القدرة على استخراجها ولكن ليس لدينا القدرة على إدخالها مرة أخرى. (******) لقد طورنا تصنيفًا لمثل هذه الخطورات المحتملة ، ووضحنا كيف أن بعضها ينتج عن التدمير الذي يصبح سهلا جدًا ، والبعض الآخر من التغييرات الخبيثة في الدوافع التي تواجه بعض الجهات الحكومية القوية أو عدد كبير من الجهات الفاعلة الضعيفة.

لقد قمنا أيضًا بفحص مجموعة متنوعة من الردود المحتملة وحدودها. كما قمنا بتتبع السبب الجذري لانكشافنا الحضاري لخاصيتين أساسيتين في النظام العالمي المعاصر: فمن ناحية ، الافتقار إلى القدرة الشرطية الوقائية لمنع الأفراد أو المجموعات الصغيرة ، بموثوقية عالية للغاية ، من تنفيذ أعمال غير قانونية للغاية ؛ ومن ناحية أخرى ، الافتقار إلى قدرة حوكمة عالمية على حل أخطر مشكلات التنسيق الدولية بشكل موثوق ، حتى عندما تحفز المصالح الوطنية الحيوية الدول ، بشكل افتراضي ، على الانشقاق. يتطلب الاستقرار العام في مواجهة نقاط الخطورة الحضارية المحتملة – في عالم يجري فيه الابتكار التكنولوجي بسرعة وفي مجالات واسعة ، وحيث توجد أعداد كبيرة من الجهات الفاعلة التي تتميز بمجموعة متنوعة من الدوافع التي لا تخفى على الإدراك البشري – أن يتم القضاء على كلتا هاتين الفجوتين الحوكميتين. وحتى يتم تحقيق ذلك ، سوف تظل البشرية عرضة لإخراج الكرة السوداء التكنولوجية.

من الواضح أن هذه الأفكار تقدم سببًا واقعيا لدعم تعزيز قدرات المراقبة وأنظمة الشرطة الوقائية وتفضيل نظام حوكمة عالمي قادر على اتخاذ إجراءات حاسمة (سواء على أساس القوة المهيمنة أحادية الجانب أو المؤسسات القوية متعددة الأطراف) . ومع ذلك ، لم نقرر ما إذا كانت هذه الأشياء مرغوبة مع مراعاة جميع الأشياء ، لأن القيام بذلك يتطلب تحليل عدد من الاعتبارات القوية الأخرى التي تقع خارج نطاق هذه الورقة.

ونظرًا لأن هدفنا الرئيسي كان وضع بعض العلامات في المشهد الاستراتيجي الكلي ، فقد ركزنا مناقشتنا على مستوى مجرد إلى حد ما ، مع تطوير المفاهيم التي يمكن أن تساعدنا في توجيه أنفسنا (فيما يتعلق بالنتائج طويلة الأجل والرغبات العالمية) بشكل مستقل ، إلى حد ما ، عن تفاصيل سياقاتنا المحلية المتنوعة.

من الناحية العملية ، إذا كان للمرء أن يبذل جهدًا لاستقرار حضارتنا في مواجهة الكرات السوداء المحتملة ، فقد يجد أنه من الحكمة التركيز منذ البدء على الحلول الجزئية والمساعي قريبة المنال. وبالتالي ، بدلاً من محاولة تكوين شرطة وقائية فعالة للغاية أو حوكمة عالمية قوية، قد يحاول المرء إصلاح مجالات معينة يبدو من المرجح أن تظهر فيها الكرات السوداء. يمكن للمرء ، على سبيل المثال ، تعزيز الإشراف على الأنشطة المتعلقة بالتكنولوجيا الحيوية من

خلال تطوير طرق أفضل لتتبع المواد والمعدات الرئيسية ، ومراقبة العلماء داخل المختبرات. يمكن للمرء أيضًا تشديد اللوائح المنظمة لطرق معرفة العملاء know-your-customer في قطاع توريد التكنولوجيا الحيوية ، وتوسيع نطاق استخدام عمليات التحقق من الخلفية لدى الموظفين العاملين في أنواع معينة من المختبرات أو المشاركين في أنواع معينة من التجارب. يمكن للمرء تحسين أنظمة المبلغين عن المخالفات ، ومحاولة رفع معايير الأمن البيولوجي على مستوى العالم. يمكن للمرء أيضًا متابعة التطور التكنولوجي التفضيلي ، على سبيل المثال ، من خلال تعزيز اتفاقية الأسلحة البيولوجية والحفاظ على الحظر العالمي للأسلحة البيولوجية. كما يمكن تشجيع هيئات التمويل ولجان الموافقة الأخلاقية على توسيع نطاق عرضها للعواقب المحتملة لخطوط عمل معينة ، مع التركيز ، ليس فقط ، على المخاطر التي يتعرض لها عمال المختبر ، وحيوانات الاختبار ، وموضوعات البحث البشرية ، ولكن أيضًا على الطرق التي قد تؤدي بها النتائج المأمولة إلى خفض مستوى الكفاءة للإرهابيين البيولوجيين في المستقبل. إن العمل الوقائي في الغالب (مثل مراقبة تفشي الأمراض ، وبناء قدرات الصحة العامة ، وتحسين أجهزة تنقية الهواء) يمكن ترقيته بشكل تفضيلي.

ومع ذلك ، أثناء السعي لتحقيق مثل هذه الأهداف المحدودة ، ينبغي للمرء أن يضع في اعتباره أن الحماية التي ستوفرها لا تغطي سوى مجموعات فرعية خاصة من السيناريوهات ، وقد تكون مؤقتة. إذا وجد المرء نفسه في وضع يسمح له بالتأثير على المعايير الكلية لقدرة الشرطة الوقائية أو قدرة الحوكمة العالمية ، يجب عليه أن يأخذ في الاعتبار أن التغييرات الأساسية في تلك المجالات قد تكون الطريقة الوحيدة لتحقيق قدرة عامة على تحقيق استقرار حضارتنا في مواجهة نقاط الخطورة التكنولوجية الناشئة.

ملاحظات

للتعليقات والمناقشة والنقد ، أنا ممتن لسونجا الصوفي Sonja Alsofi ، ستيوارت أرمسترونج Stuart Armstrong ، أندرو سنايدر بيتي Andrew Snyder-Beattie ، كريس أندرسون Chris Anderson ، نِك بيكستيد Nick Beckstead ، مايلز بروندج Miles Brundage ، بن بوكانان Ben Buchanan ، أُوِن

كوتون–بارات Owen Cotton–Barratt ، نيل باورمان Niel Bowerman ، بول كريستيانو Paul Christiano ، ألان دافو Allan Dafoe ، جيف دينج Jeff Ding ، إريك دريكسلر Eric Drexler ، بيتر إكِرسلي Peter Eckersley ، أوِين إيفانز Owain Evans ، توماس هومر–ديكسون Thomas Homer–Dixon ، توماس إنجليسبي Thomas Inglesby ، جون ليزلي John Lesilie ، جريجوري لويس Gregory Lewis ، ماتيس ماس Matthijs Maas ، جيسون ماثيني Jason Matheny ، مايكل مونتاج Michael Montague ، لوك مويلهاوزر Luke Muehlhauser ، توبي أورد Toby Ord ، بن بيس Ben Pace ، ريتشارد ري Richard Re ، أندرس ساندبرج Anders Sandberg ، جوليان سافوليسكو Julian Savulescu ، ستيفان شوبرت Stefan Schubert ، كارل شولمان Carl Shulman ، تانيا سينغ Tanya Singh ، هيلين تونر Helen Toner ، ولجماهير الكثير من ورش العمل والمحاضرات حيث جرى تقديم نسخ سابقة من هذا العمل) ، ولثلاثة حكام مجهولين ؛ وأشكر كاريك فلين Carrick Flynn و كريستوفر جالياس Christopher Galias ، بن جارفينكل Ben Garfinkel ، روز حدشر Rose Hadshar للمساعدة في إعداد المخطوطة الأولى والكثير من الاقتراحات المفيدة.

تلقى هذا العمل تمويلًا من مجلس البحوث الأوروبي (ERC) في إطار برنامج البحث والابتكار التابع للاتحاد الأوروبي (Horizon 2020) (اتفاقية المنحة رقم ٦٦٩٧٥١) .

١– من الواضح أن استعارة الجرة لها محددات مهمة. سوف نناقش بعضها لاحقًا.

٢– من الصعب تقييم التأثير الخالص على أحوال الحيوانات غير البشرية. على وجه الخصوص ، تنطوي الزراعة الصناعية الحديثة على سوء معاملة أعداد كبيرة من الحيوانات.

٣– ومع ذلك ، هناك أمثلة على "الثقافات" أو السكان المحليين الذين قد يكون سبب وفاتهم (جزئيًا على الأقل) ممارساتهم التكنولوجية الخاصة ، مثل جزيرة الفصح Easter Island (Rapa Nui) شعب الهنود الحمر الأجداد في جرين لاند Mesa Varde (المكسيك الجديدة Anasazi) ، الذين ، وفقًا لـ Diamond 2005 ، قطعوا غاباتهم الخاصة ثم عانوا انهيارا بيئيا.

٤– على أي حال ، يمكن العثور على أمثلة في الأنواع الأخرى ، إذا أخذنا في الاعتبار التكيفات التطورية باعتبارها اختراعات. على سبيل المثال ، هناك مؤلفات تستكشف كيف تكون نهاية أعمال التطور ، والتي تؤدي إلى انقراض نوع أو مجموعة ما ، قد تنجم عن تغييرات تطورية مفيدة مثل تلك التي ينطوي عليها التخصص (حيث قد يؤدي التكيف مع جانب ضيق إلى خسارة لا رجعة فيها للسمات اللازمة للبقاء على قيد الحياة في نطاق أوسع من البيئات) (Day et al. 2016) ، وظهور النظم الاجتماعية الفطرية (مثل العناكب الاجتماعية) (Aviles and Purcell, 2012) ، أو التحول إلى الأنانية (على سبيل المثال بين النباتات المزهرة الأنواع التي تنتقل من التهجين إلى الإخصاب الذاتي) (Igic and Busch, 2013) .

٥- على الرغم من أن معظم العلماء المشاركين في المشروع يفضلون المقترحات مثل خطة باروخ Baruch Plan ، التي كانت ستضع الطاقة النووية تحت السيطرة الدولية ، وبقي لديهم القليل من سلطة اتخاذ القرار في هذه المرحلة.

٦- مجازيا بالطبع. ولكن يمكن القول بلغة الاستعارة أنه ينبغي أن يكون هناك أكثر من إصبع مرتجف على كل جانب ، بالنظر إلى القيادة والسيطرة المفوضة على نطاق واسع (Ellsberg, 2017; Schlosser, 2013) .

٧- ومع ذلك ، داخل دولة معينة ، ربما يكون عدد الجهات الفاعلة التي لديها سلطة شن هجمات نووية كبيرا جدًا. إلسبرج (٢٠١٧) يزعم أنه ، على الأقل على مدى جزء كبير من الحرب الباردة ، تمت مناقشة سلطة إطلاق أسلحة نووية متعددة لتتدرج تحت تسلسل القيادة الأمريكية. كان عدد الضباط الذين كانت لديهم القدرة المادية على إطلاق أسلحة نووية بالضرورة كبيرا ، على الرغم من أنها ليست السلطة للقيام بذلك. في الاتحاد السوفيتي ، في مرحلة ما أثناء الانقلاب على ميخائيل جورباتشوف في أغسطس ١٩٩١ ، كانت جميع ("الحقائب النووية") Chegets الثلاثة في الاتحاد السوفيتي في أيدي قادة الانقلاب (سوكوسكي و ترترايس ، ٢٠١٣ ؛ ستيفنسون ، ٢٠٠٨) .

٨- إن "خطر المعلومات" هو الخطر الناشئ عن نشر معلومات صحيحة ، على سبيل المثال لأن المعلومات يمكن أن تمكن بعض العملاء من إحداث ضرر. بوستروم (2011) يناقش أخطار المعلومات بشكل أكثر عمومية.

٩- قد يقولون بأن الانفتاح سيكون مفيدًا لأنه سوف يسمح لمزيد من الناس بالعمل على التدابير المضادة (راجع النقاش حول اكتساب وظيفة العمل على فيروسات الإنفلونزا ؛ Duprex et al., 2015; Fauci et al., 2011; Sharp, 2005) . قد يقولون أيضًا بأنه ما دامت الحكومة مستمرة في تبرير الإجراءات الوحشية بالإشارة إلى معلومات سرية ، فإنها بذلك ستكون غير مسؤولة بشكل خطير أمام مواطنيها. اعتقاد مماثل دفع المجلة الأمريكية The Progressive's decision في أواخر سبعينيات القرن العشرين إلى نشر أسرار حول القنبلة الهيدروجينية ، حتى في مواجهة الطعن القانوني من قبل وزارة الطاقة في الولايات المتحدة. مؤلف الفقرة ، هوارد مورلاند ، كتب: "السرية ذاتها ، لا سيما سلطة عدد قليل من "الخبراء" المعينين لإعلان بعض الموضوعات خارج الحدود ، تساهم في مناخ سياسي حيث يمكن للمؤسسة النووية أن تمارس الأعمال كالمعتاد ، بما يحمي ويديم إنتاج أسلحة الرعب هذه" Morland (1979, p. 3) .

١٠- بشكل عام ، في الحالات التي يكون فيها لكل من الجهات الفاعلة المتعددة احتمالية مستقلة لاتخاذ إجراء من جانب واحد ، تميل احتمالية اتخاذ الإجراء إلى احتمال واحد هو أن يزيد عدد الجهات الفاعلة. عندما تنشأ هذه الظاهرة عند الجهات الفاعلة ذات الأهداف المشتركة ولكن أحكامها متضاربة ، بسبب العشوائية في الأدلة التي يتعرضون لها أو المنطق الذي يتبعونه ، هناك ينشأ "بلاء أحادي الجانب" (Bostrom, 2016) . البلاء

يدل على أنه حتى القرار غير الحكيم للغاية ، مثل قرار نشر تصميمات الأسلحة النووية ، من المرجح أن يتم إذا كان هناك عدد كافٍ من الجهات الفاعلة في وضع يسمح لها باتخاذه من جانب واحد.

١١- الكثير من هذه الدوافع ذاتها واضحة اليوم بين هاكرز القبعات السوداء الذين ينفذون هجمات إلكترونية ضارة. على سبيل المثال ، كطريقة للابتزاز ، لقد أثبت بعض المتسللين المجهولين رغبتهم في شل قدرات المدن على تقديم خدمات حيوية لسكانها (Blinder and Perlroth, 2018) . يبدو أن الدوافع وراء الهجمات الإلكترونية الضارة اقتصاديًا تشمل أيضًا كل من الإيديولوجيا السياسية والفضول. حيث أن هجمات الإنترنت المعاصر أقل تدميراً بشكل كبير من هجمات الأسلحة النووية ، إن مجموعة الجهات الفاعلة التي قد تكون على استعداد لاستخدام الأسلحة النووية أو التهديد باستخدامها هو بالتأكيد أصغر بكثير من مجموعة الجهات الفاعلة المستعدة للانخراط في قرصنة ضارة. ومع ذلك ، قد تكون العوامل الاجتماعية والنفسية ذات الصلة بكلتا الحالتين متشابهة.

١٢- يختلف هذا المفهوم عن مفهوم الفوضى العالمية في مجال العلاقات الدولية. المفهوم الحالي يؤكد أن الفوضى تعد مسألة درجة ، ويقصد بها أن تكون محايدة نسبيًا كما هو الحال بين المدارس الفكرية المختلفة في العلاقات الدولية. راجع (Lechner, 2017) . والأهم من ذلك ، أنه يشمل الافتقار إلى الحوكمة ، ليس فقط "في القمة" ولكن أيضًا "في القاع" . وهذا يعني أن الحالة الافتراضية شبه الفوضوية تشير إلى حقيقة أنه في النظام العالمي الحالي ، لا توجد فقط درجة من الفوضى على المستوى الدولي ، بسبب نقص الحوكمة العالمية أو غيرها من الوسائل الفعالة بالكامل لتقييد تصرفات أي دولة وحل مشكلات التنسيق العالمي ، ولكن هناك أيضًا درجة من الفوضى على مستوى الأفراد (والجهات الفاعلة الأخرى التابعة للدولة) في ذلك ، حتى الدول عالية التنظيم الوظيفي تفتقر حاليًا إلى القدرة على تنظيم تصرفات هؤلاء الفاعلين الصغار. على سبيل المثال ، على الرغم من سعي الكثير من الدول إلى منع الاغتصاب والقتل داخل أراضيها ، لا يزال الاغتصاب والقتل يحدثان بوتيرة غير صفرية. عواقب هذا هي أن هذه الدرجة من الفوضى في القاع يمكن أن تتضخم إلى حد كبير إذا حصل الأفراد على قدرات تدميرية أكبر بكثير.

١٣- للمقارنة ، عدد القتلى ١٥ في المائة من عدد سكان العالم الحاليين ، وهذا يعد أكثر من ضعف الآثار المجمعة للحرب العالمية الأولى ، والإنفلونزا الإسبانية ، والحرب العالمية الثانية كنسبة مئوية من سكان العالم (والفرق أكبر من حيث القيمة المطلقة) . يعتبر الانخفاض بنسبة ٥٠ في المائة في إجمالي الناتج العالمي أكبر انخفاض في التاريخ المسجل. خلال فترة الكساد الكبير ، على سبيل المثال ، انخفض إجمالي الناتج العالمي بما يقدر بنحو ١٥ في المائة أو أقل ، وتم استردادها في الغالب في غضون بضع سنوات (على الرغم من أن بعض النماذج تشير إلى أن ذلك كان له أيضًا تأثير محبط طويل الأمد على التجارة مما أضعف الاقتصاد العالمي بشكل مزمن) (Bolt et al., 2018) ؛ (Crafts and Fearon, 2010) .

١٤- تركز هذه الورقة على نقاط الخطورة التكنولوجية المحتملة. ويمكن أن تكون هناك أيضًا نقاط خطورة طبيعية محتملة تنشأ بشكل مستقل عن تقدم الحضارة الإنسانية ، مثل وابل النيازك العنيف الذي سيؤثر على كوكبنا في تاريخ ما في المستقبل. يمكن تحقيق الاستقرار لبعض نقاط الخطورة الطبيعية المحتملة بمجرد أن يتجاوز مستوى قدرتنا التكنولوجية عتبة معينة (مثل القدرة على تشتيت النيازك) . من المعقول أن يكون خطر نقاط الخطورة التكنولوجية المحتملة أكبر من مخاطر نقاط الخطورة الطبيعية المحتملة ، على الرغم من أن الحجة لهذا الأمر أقل وضوحًا مع تقليل شدة الدمار الحضاري مما سيكون عليه الحال إذا جرى هذا التقليل على كارثة وجودية (Bostrom, 2013; Bostrom and Cirkovic, 2011) . التحفظ (الكبير) على هذا الادعاء (الخاص بالخطورة التكنولوجية المهيمنة) هو أنه يفترض مسبقًا أن العالم لا ينزف القيمة المحتملة بمعدل كبير. إذا قمنا بدلاً من ذلك بتقييم الأشياء من منظور يصح فيه ما يمكن أن نطلق عليه فرضية العالم النازف ، عندئذٍ قد يكون الخراب الافتراضي الناشئ عن الطبيعة (أي غير البشرية) مهيمنا على المعادلة. ويمكن تقديم فرضية العالم النازف ، على سبيل المثال: (١) المقيِّم يهتم كثيرًا بالأشخاص الموجودين (بما في ذلك الذات والعائلة) وهم بطبيعة الحال يموتون بمعدل كبير (على سبيل المثال ، من الشيخوخة) ، وبالتالي فقدان القدرة على الاستمرار في الاستمتاع بحياتهم وفرصة الوصول إلى مستويات أعلى بكثير من الرفاهية التي قد تصبح ممكنة عند النضج التكنولوجي ؛ (٢) المقيِّم يهتم كثيرًا بتجنب المعاناة التي يمكن تجنبها باستخدام التكنولوجيا الأكثر تقدمًا لكن هذا يحدث حاليًا ، مما يؤدي إلى تراكم الفوضى ؛ (٣) هناك بعض المعدلات الجوهرية الخارجية للتدمير الحضاري (مثل الكوارث الطبيعية ، وإنهاء المحاكاة العشوائية غير المرتبطة بأنشطتنا (Bostrom, 2003) . وبينما نسمح بمرور الوقت ، فإننا نتحمل خطرًا تراكميًا لتدميرنا قبل الوصول إلى الحد الأقصى من الإمكانات التكنولوجية : (٤) هناك طرق لتستخدم في الفيزياء لا نفهمها حاليا بشكل جيد لبدء عمليات نمو سريع لخلق القيمة (مثل إنشاء سلسلة متسارعة من الأكوان الصغيرة التي سيكون سكانها سعداء للغاية) ، ويهتم المقيّم بمثل هذا الخلق بطريقة المقياس الحساس ؛ و (٥) الدوائر الانتخابية فائقة الذكاء الأخرى ، التي هي في وضع يمكنها من التأثير بشكل كبير على الأشياء التي يهتم بها المقيِّم ، ونفاد صبرها لوصولنا إلى مستوى معين من التقدم ، لكن القيمة التي يضعونها على ذلك تتحلل بسرعة بمرور الوقت.

١٥- يمكن أن يظل العالم عرضة للخطورة المحتملة بعد تراجع التكنولوجيا العميقة ، على سبيل المثال إذا بقي الكثير من الأسلحة النووية سابقة التصنيع ، حتى بعد أن تتراجع الحضارة ، فوق نقطة عجزها عن تصنيع أسلحة نووية جديدة.

١٦- من المهم بالنسبة لسيناريو "الأسلحة النووية السهلة" الأصلي أن يتطلب كل استخدام نووي جهود فرد واحد أو جهود مجموعة صغيرة. مع أنه قد يتطلب جهودًا مشتركة من المئات من الجهات الفاعلة لتدمير الحضارة في ذلك السيناريو - رغم كل شيء ، يختلف تدمير مدينة أو منطقة حضرية واحدة عن تدمير الحضارة - هذه المئات من الجهات الفاعلة لا تحتاج إلى التنسيق. وهذا يسمح للانفجار المروع أن يلعب دوره.

١٧- يقدم بوم وآخرون (2018) ,.Baum, et all قائمة محدثة بالحوادث النووية والمناسبات التي تم فيها التفكير في استخدام الأسلحة النووية. ويقدم ساجان Sagan عام (١٩٩٥) وصفًا أكثر شمولاً للممارسات الخطيرة خلال الحرب الباردة. كما قام شلوسر Schlosser عام (٢٠١٣) بفحص الحوادث غير النووية ، مع التركيز بشكل خاص على حادثة واحدة أدت إلى تفجير غير نووي لصاروخ تيتان ٢ العابر للقارات Titan-II ICBM .

١٨- على الرغم من وجود عدد قليل من المحاولات العلمية لتقييم درجة الحظ المتضمَّنة في تجنب هذه النتيجة ، حيث يضع أحد التقديرات الحديثة ، بالاعتماد على مجموعة بيانات لحالات وشيكة الوقوع ، احتمال تجنب الولايات المتحدة والاتحاد السوفيتي للحرب النووية بنسبة أقل من ٥٠ في المائة (Lundgren, 2013). وهذا يتماشى مع آراء بعض المسؤولين الذين لديهم معرفة من الداخل بالأزمات النووية ، مثل الرئيس جون كينيدي ، الذي أعرب عن اعتقاده ، بعد فوات الأوان ، أن أزمة الصواريخ الكوبية كانت لديها الفرصة بين واحد من اثنين وواحد من ثلاثة لتؤدي إلى حرب نووية. ومع ذلك ، فإن عددًا من علماء الأمن الدوليين البارزين ، مثل كينيث والتز Kenneth Waltz وجون مولر John Mueller ، يرون أن احتمالية نشوب حرب نووية كانت منخفضة للغاية وبصورة مستمرة (Mueller, 2009; Sagan and Waltz, 2012) .

١٩- ربما يُعتقد بطريق الخطأ ، بحسب قائد سابق في أسطول المحيط الهادئ الأمريكي ، أنه كانت هناك فترة خلال الحرب الباردة أصبحت فيها المراقبة المضادة للغواصات فعالة للغاية: "يمكن للولايات المتحدة أن تحدد من خلال رقم الهيكل هوية الغواصات السوفيتية ، ولذلك كان باستطاعتنا إجراء إحصاء لقوتها ومعرفة مكانها بالضبط. في الميناء أو في البحر. ... لذلك شعرتُ بالارتياح لأننا حصلنا على القدرة على القيام بشيء خطير للغاية لقوة SSBN السوفيتية من خلال ملاحظة قصيرة جدًا وسط أي مجموعة من الظروف تقريبًا." (quoted in Ford and Rosenberg, 2005, p. 399) .

٢٠- في الواقع ، التطورات في الاستشعار عن بعد ، ومعالجة البيانات ، والذكاء الاصطناعي ، والطائرات بدون طيار ، وأنظمة التوصيل النووية تهدد الآن بتقويض الردع النووي ، خاصة بالنسبة للدول ذات الحجم الصغير نسبيًا والترسانات النووية غير المتطورة (Lieber and Press, 2017) .

٢١- لم نكن في الحقيقة خارج نطاق الضرر ، بالطبع: التساقط الإشعاعي سوف يؤثر على الحلفاء وعلى الوطن الى حد ما؛ التداعيات الاقتصادية من شأنها أن تنشر الخراب في الأسواق وتؤدي إلى كساد عالمي. ومع ذلك ، سيكون من الأفضل كثيرًا أن نكون هدفًا للهجوم (خاصة إذا وضعنا جانبًا الشتاء النووي) .

٢٢- الاحتمال الآخر هو أنه ستكون هناك مكاسب سياسية ، مثل زيادة القدرة على ممارسة الإكراه النووي ضد أطراف ثالثة بعد أن أبدوا استعدادهم لاستخدام الأسلحة النووية.

23- Brooks (1999); Gartzke (2007); Gartzke and Rohner (2011).

للاطلاع على وجهة نظر مناقضة انظر ليبرمان ١٩٩٣.

٢٤- ربما لم يكن من الممكن أن تنشأ الحضارة البشرية لو كان مناخ الأرض حساسًا لثاني أكسيد الكربون ، حيث كانت مستويات ثاني أكسيد الكربون في الماضي (٤٠٠٠ جزء في المليون خلال العصر الكمبري / عصر الحيتان مقارنة بحوالي ٤١٠ جزء في المليون اليوم) من المفترض أن يكون قد تعطل تطور الحياة المعقدة بشكل خطير. قد يتضمن الواقع المغاير المنفصل قليلا ، بدلاً من ذلك ، نوعاً من المركبات التي لا توجد بكميات كبيرة في الطبيعة ولكن تنتجها الحضارة الإنسانية ، مثل مركبات الكربون الكلورفلوري. تم التخلص التدريجي من مركبات الكربون الكلورفلوري عبر بروتوكول مونتريال بسبب تأثيرها المدمر على طبقة الأوزون ، ولكنها أيضًا غازات دفيئة شديدة الفعالية على أساس القياس في كل كيلوجرام. لذلك يمكننا أن نفكر في الواقع المضاد الذي كانت فيه مركبات الكربون الكلورفلوري مفيدة صناعيا على نطاق أكبر بكثير مما كانت عليه ، ولكن مع مراعاة الآثار التراكمية الدرامية المتأخرة على المناخ العالمي.

٢٥- ينتهي التقرير الذي أمر أوبنهايمر بإعداده بالكلمات التالية: "يمكن للمرء أن يستنتج أن الحجج الواردة في هذه الورقة تجعل من غير المعقول توقع أن ينتشر تفاعل N + N بصورة دعائية. وانتشارا الدعاية غير المحدود أقل احتمالا. ومع ذلك ، فإن تعقيد الحجة وغياب الأساس التجريبي المُرضي يجعل المزيد من العمل حول هذا الموضوع مرغوبًا للغاية" (Konopinski et al. 1946) .

٢٦- يمكن النظر إلى الخطورة المحتملة من النوع صفر على أنها الحالة المحدِّدة للنوع ١ : فهي تشير إلى خطورة محتملة لا تتطلب أية جهات فاعلة سيئة النية من أجل إنتاج الدمار الحضاري – يكفي فقط وجود الجهات المسؤولة الفاعلة عادةً الذين هم على استعداد للمضي قدما في استخدام التكنولوجيا إلى ما بعد المستوى المعتاد من الأمن الذي يجري إعطاؤه للتكنولوجيا الجديدة.

٢٧- وإذا كانت ١٠ سنوات ، فلماذا لا تكون دائمة.

٢٨- في الواقع ، يشير تقرير ألبرت سبير Albert Speer ، وزير التسليح الألماني ، إلى أن فيرنر هايزنبرج ناقش إمكانية حدوث تفاعل متسلسل مع هتلر وأن هذه الإمكانية ربما أضعفت حماس هتلر لمتابعة القنبلة. (Rhodes, 1986) .

٢٩- إن نسخة حقيقية من هذا النوع من الخطورة المحتملة من النوع ٢ أ ، والتي تواجه فيها الجهات الفاعلة الرئيسية دوافع استراتيجية لاتخاذ الإجراءات التي تخلق مخاطر غير مرغوبة للحضارة ، يمكن أن تنشأ في سياق السباق نحو تطوير الذكاء الخارق للآلة. في الظروف غير المواتية ، يمكن أن تقدم الديناميكيات التنافسية

مطوِّرا رائدا مع الاختيار بين إطلاق الذكاء الاصطناعي الخاص به قبل أن يصبح آمنًا أو التخلي عن زمام المبادرة لأحد المطورين الآخرين الذين هم على استعداد لفعل مخاطر أكبر (Armstrong et al., 2016).

٣٠- لمناقشة كيف يوازن المخطِّط العقلاني بين نمو الاستهلاك والسلامة في نماذج مختلفة حيث يحمل الابتكار المحفز للنمو أيضًا مخاطر إدخال ابتكارات تقلل من العمر. انظر (Jones, 2016) .

٣١- حتى سيناريو "الخنق المفاجئ" قد تربكه مشكلات التنسيق ، وإن كانت بدرجة أقل من "القلعة برافو / اختبار الثالوث" . قد يكون للأشخاص ، الذين يتخذون قرارًا بشأن مخصصات تمويل العلوم ، أولويات مختلفة عن الجمهور الذي يقدم التمويل. قد يضعون ، على سبيل المثال ، قيمة أعلى لإرضاء الفضول الفكري بالنسبة إلى القيمة التي يعطونها لعملية إبقاء المخاطر منخفضة ، وتوفير فوائد مادية على المدى القريب للجماهير. يمكن أن تؤدي مشكلات العامل الرئيسي عندئذٍ إلى تمويل أكبر لمسرعات الجسيمات مما قد تحصل عليه مثل هذه التجارب في غياب مشكلات التنسيق. سباقات الهيبة / الهيمنة بين الأمم – والتي قد يُنظر إليها جزئيًا على أنها فشل آخر في التنسيق – قد يمثل أيضًا محركًا لتمويل العلوم الأساسية بشكل عام وفيزياء الطاقة العالية بشكل خاص.

٣٢- قد يفضل الناس ، الذين هم على قيد الحياة وقت حدوث الدمار الشامل ، لو أنه حدث قبل ذلك ، قبل ولادتهم ، لكي ينتهي كل شيء ولا يتأثروا. يبدو أن تفضيلاتهم تواجه مشكلة غير متعلقة بالهوية ، حيث أنه لو حدث تدمير حضاري قبل أن يكونوا في مجال التصور فلن يكونوا على يقين من وجودهم (Parfit, 1987).

٣٣- وعلى نطاق أوسع ، هناك العديد من التحسينات في أدوات وتقنيات التكنولوجيا الحيوية ، والتي تسهل على هواة القرصنة البيولوجيين إنجاز ما كان لا يمكن القيام به في السابق إلا عن طريق مختبرات البحوث المهنية المزودة بالموارد الجيدة ، والتي تدخل في إطار الشك من هذا المنظور. من المشكوك فيه للغاية ما إذا كانت فوائد القرصنة البيولوجية DIY (النباتات المنزلية المتوهجة في الظلام؟) تستحق تكاثر القدرة على تحويل الهندسة الحيوية إلى أغراض يحتمل أن تكون خطيرة أو ذات أغراض ضارة لمجموعة موسعة من الجهات الفاعلة غير الخاضعة للمساءلة نسبيًا.

٣٤- الحجة المضادة "إذا لم أطورها ، فسيقوم شخص آخر بذلك ؛ لذلك ينبغي أن أفعل ذلك أيضًا" تميل إلى التغاضي عن حقيقة أن عالِمًا أو مطورًا معينًا له على الأقل بعض التأثير الهامشي على التوقيت المتوقع للاكتشاف الجديد ، وإذا كان هذا هو الحال حقا أنه لا يمكن لجهود العالِم أن تحدث فرقًا في وقت الاكتشاف أو الاختراع ، إذن ، يبدو أن الجهود مضيعة للوقت والموارد ، وينبغي إيقافه لهذا السبب. إن تحولا صغيرا نسبيا في نوع من القدرة التكنولوجية وجعله متاحًا (على سبيل المثال ، خلال شهر واحد) قد يكون مهمًا في بعض السيناريوهات (مثل: ما إذا كانت التكنولوجيا الخطيرة تفرض درجة كبيرة من الخطورة كل شهر حتى يتم تطوير ونشر دفاعات فعالة) .

٣٥- نعني بعبارة "النضج التكنولوجي" بلوغ القدرات التي توفير مستوى من الإنتاجية الاقتصادية والتحكم في الطبيعة قريبا من الحد الأقصى الذي يمكن تحقيقه عمليًا (في امتلاء / تشبع الوقت) (Bostrom, 2013) .

٣٦- ازدادت أهمية التحكم في الوصول إلى العلوم البيولوجية منذ حادث الجمرة الخبيثة 2001 "Amerithrax" في الولايات المتحدة ، حيث تم إلزام المؤسسات التي تتعامل مع مسببات الأمراض الخطيرة بتقييم مدى ملاءمتها للموظفين الذين سوف يكون لديهم إمكانية الوصول إليها ، والتي يجري فحصها أيضًا من قِبل الوكالات الفيدرالية (Federal Select Agent Program, 2017) ، والتوجهات المشابهة الموصى بها للبلدان التي تقوم بتطوير بنيتها التحتية للأمن البيولوجي (مركز للأمن البيولوجي والتأهب البيولوجي ، ٢٠١٧) . يعاني النظام الحالي من عيبيْن: أولاً ، لا يوجد تنسيق عالمي. لذا يمكن للأطراف السيئة أن "تتجول" بحثًا عن بيئات تنظيمية أكثر مرونة؛ ثانيًا ، يظل التركيز على الوصول إلى المواد البيولوجية (مثل عينات من كائنات دقيقة معينة) ، في حين تكون المعلومات والتكنولوجيا البيولوجية هي على نحو متزايد ، الموضوع الرئيسي للاهتمام الأمني (Lewis et al., 2019) .

٣٧- يبدو أن قيمة X التي تقل كثيراً عن ١ في المائة متسقة مع قلة ما يقدمه معظم الناس للجمعيات الخيرية العالمية. من ممكن القول ، مع ذلك ، أن التمييز بين الفعل والإغفال سوف يجعل الناس يرغبون في قبول تضحية شخصية من أجل عدم المساهمة في عالم سيء أكبر بكثير مما لو كانوا سيفعلون من أجل المساهمة في الصالح العام.

٣٨- لاحظ ، مع ذلك ، أن التحول الإيجابي في توزيع الأفضليات – حتى لو لم يكن كافيا لتفادي وقوع كارثة بمجرد عدم اختيار بعض الجهات الفاعلة الفردية للخيار المدمر – يمكن أن يكون له تأثيرات مهمة غير مباشرة. على سبيل المثال ، إذا كان هناك عدد كبير من الناس أصبح أكثر ميلا للخير قليلا ، وهذا قد يغير المجتمع إلى توازن أكثر تعاونًا من شأنه أن يدعم أساليب استقرار أقوى تقوم على الحوكمة مثل تلك التي نناقشها أدناه ، راجع (moral enhancements, Persson and Savolescu, 2012) .

٣٩- على المستوى العالمي ، نجد توليفة من مخططات التصنيف الوطني وأنظمة التحكم في المعلومات. وهي مصممة بشكل عام لحماية الأسرار العسكرية والاستخبارية ، أو لمنع حقائق محرجة حول المطلعين على النظام من العرض على الجمهور ، ليس لتنظيم انتشار أفكار العلم أو التكنولوجيا. هناك بعض الاستثناءات ، لا سيما في حالة المعلومات الفنية التي لها تأثير مباشر على الأمن القومي. على سبيل المثال ، قانون سرية الاختراع لعام ١٩٥١ في الولايات المتحدة ، الذي يمنح وكالات الدفاع سلطة إيقاف منح براءة اختراع وتأمر بالحفاظ على سرية اختراع ؛ على الرغم من أن المخترِع الذي يمتنع عن السعي للحصول على حماية براءات الاختراع لا يخضع لهذه القيود (Parker and Jacobs 2003) . تخضع الاختراعات النووية لبند "السر المولود" من قانون الطاقة الذرية لعام ١٩٤٦ ، والذي يوضح جميع المعلومات المتعلقة بالتصميم ، والتطوير ، وتصنيع

الأسلحة النووية – بغض النظر عن المنشأ – تعد سرِّية ما لم يتم رفع السرية عنها رسميًا (Parker and Jacobs 2003) . كما تم استخدام أدوات قانونية أخرى ، مثل ضوابط التصدير ، في محاولات لوقف تدفق المعلومات العلمية. كما تقدم الجهود (غير الناجحة) التي يبذلها عدد من الوكالات الحكومية الأمريكية لمنع نشر واستخدام بروتوكولات التشفير القوية ، التي تم تطويرها في السبعينيات والثمانينيات من القرن الماضي ، مثالًا بارزًا (Banisar, 1999) . لقد تم تجريب الرقابة الذاتية الطوعية من قِبل المجتمع العلمي في مناسبات نادرة جدا. وحقق ليو تسيلارد بعض النجاحات الجزئية في إقناع زملائه الفيزيائيين بالامتناع عن النشر عن جوانب من الانشطار النووي (قبل بدء مشروع مانهاتن وبدء السرية الرسمية) ، على الرغم من أنه واجه معارضة من بعض العلماء الذين أرادوا ظهور عملهم الخاص في المجلات ، أو الذين شعروا أن الانفتاح قيمة مقدسة في العلم. في الآونة الأخيرة ، كانت هناك بعض المحاولات للرقابة الذاتية العلمية فيما يتعلق بأبحاث أنفلونزا الطيور (Gronvall, 2013) . في هذه الحالة ، ربما لم تكن الجهود غير فعالة فحسب ، بل كانت لها نتائج عكسية ، حيث إن الجدل الذي أثاره النقاش المفتوح حول ما إذا كان ينبغي نشر نتائج معينة قد جذب المزيد من الانتباه لتلك النتائج أكثر مما لو تم نشرها دون معارضة – ما يسمى بـ "تأثير ستراايسند Streisand effect" . يبدو أن محاولات الرقابة الذاتية العلمية كانت ، بشكل عام ، فاترة إلى حد ما وغير فعالة. (أقول يبدو ، بسبب كيفية تطور الأمور في الحلقات المعروفة للجمهور حيث جرت محاولات الرقابة. ولكن المحاولات الناجحة حقًا لحجب المعلومات العلمية لن تظهر بالضرورة في السجل العام.) . حتى لو تمكن عدد قليل من محرري المجلات من الاتفاق على معايير لكيفية التعامل مع الأوراق التي تشكل مخاطر على المعلومات ، فلا شيء يمنع مؤلف محبط من إرسال مخطوطته إلى مجلة أخرى ذات معايير أقل أو نشرها على صفحة الإنترنت الشخصية. معظم المجتمعات العلمية ليس لديها ثقافة ، ولا دوافع ، ولا خبرة في الأمن وتقييم المخاطر ، ولا آليات الإنفاذ المؤسسي التي ستكون مطلوبة للتعامل بشكل فعال مع مخاطر المعلومات. الروح العلمية بالأحرى: تقتضي إخراج كل كرة من الجرة بأسرع ما يمكن وكشفها على الفور للجميع في أنحاء العالم ؛ بقدر ما يحدث هذا ، بقدر إحراز المزيد من التقدم ؛ وكلما زادت مساهمتك في هذا ، كلما كنتَ عالِمًا أفضل. إمكانية وجود الكرة السوداء لا تدخل في المعادلة.

٤٠– وعلى أية حال ، ليس من الواضح ما إذا كنا نريد حقاً أن نكون أكثر حذرا بشكل عام. قد يكون من المرغوب به (من وجهات نظر تقييمية مختلفة) لتشجيع مزيد من الحذر على وجه التحديد في المواقف حيث يمكن أن تكون هناك سلبيات عالمية شديدة. إلا أن التحذيرات الدافعة إلى ممارسة الحذر الطوعي وضبط النفس في هذه الأسباب قد لا تكون فعالة للغاية إذا كان سبب المخاطرة المفرط معياريا يعد مشكلة تنسيق: يكتسب المجازِف بعض المنفعة الخاصة (مثل الربح أو المكانة) مع توليد عوامل خارجية للمخاطر العالمية. في مثل هذه الحالات ، وبناء عليها ، قد يتطلب الحل تقوية قدرة الحوكمة العالمية.

٤١- من الممكن أيضاً أن تزيد أعمال تقييم المخاطر من مستوى الخطورة من خلال توليد مخاطر المعلومات . (Bostrom, 2011)

٤٢- الاسم الذي يبدو أورويلِيًّا (نسبة إلى Orwell) هو بالطبع مقصود لتذكيرنا بمجموعة الطرق كاملة التي يمكن من خلالها تطبيق مثل هذا النظام.

٤٣- تفاصيل التنفيذ هي للتوضيح فقط؛ على سبيل المثال ، يمكن تقديم وظيفية مماثلة من خلال نظارات الواقع المختلط بدلاً من القلادة. يمكن تصميم إصدارات من الجهاز بحيث تقدم الكثير من الفوائد للمستخدم إلى جانب وظيفة المراقبة التي تؤديها. من الناحية النظرية ، يمكن أن تكون بعض عمليات المراقبة نابعة من المصادر الجماعية: عندما يكتشف الذكاء الاصطناعي نشاطًا مشبوهًا ، تكون تغذية الفيديو قد تم محو هويتها وإرسالها إلى ١٠٠ مواطن بشكل عشوائي ، واجبهم هو مشاهدة الخلاصة والتصويّت على ما إذا كان يتطلب مزيدًا من التحقيق ؛ إذا اعتقد ما لا يقل عن ١٠ في المائة منهم أنه كذلك ، فإن النسخة (غير مجهول الهوية) يتم إرسالها إلى السلطات.

٤٤. أمثلة على "وسائل الملاءمة" التي من شأنها أن تؤدي بشكل معقول إلى مزيد من التدخل بالمراقبة تشمل أنواعًا مختلفة من تطبيقات المستهلك والمراقبة المفيدة أو المربحة اقتصاديًا (على سبيل المثال ، لاستهداف الإعلانات ، وتمييز السعر ، إلخ) ؛ القدرة على منع الأشياء المختلفة التي تسبب غضبًا عامًا ، مثل إساءة معاملة الأطفال أو الإرهاب على نطاق صغير ؛ وخاصة بالنسبة للأنظمة الاستبدادية ، والقدرة على قمع المعارضة السياسية.

٤٥- إن وجود نسخة مخففة من السياسة قد يمنع هؤلاء المشتبه بهم الضعفاء من الوصول إلى المعدات والمواد اللازمة لإنتاج التأثير المدمر. والمدى الذي قد يكون كافيا يعتمد على تفاصيل السيناريو.

٤٦. قد يحل التنفيذ الجزئي للمراقبة عالية التقنية محل الحبس في هذا السيناريو ، حيث لا يوجد سوى أولئك الموجودين في قائمة طويلة من "الأفراد ذوي الاهتمام الشديد" المطلوب ارتداؤهم قلادات الحرية.

٤٧. بالطبع ، من الممكن من الناحية النظرية أن يكون من شأن أي من هذه المعالجات أن يرفع ، بدلاً من أن يخفض ، الخطورة الحضارية المحتملة الأقل تعرضًا للكرة السوداء المحتملة ، على سبيل المثال ، إذا كان التنسيق العالمي المناسب جعل الشرطة الوطنية الفعالة للغاية أقل احتمالا ، أو العكس ، فإن طبيعة الأنظمة التي تميل إلى الظهور ، في ظل ظروف الشرطة الوقائية الأقوى أو الحوكمة العالمية ، تختلف أيضًا عن أولئك الذين يعيشون في الوضع الراهن بطرق من شأنها أن تزيد أو تقلل بعض نقاط الخطورة الحضارية المحتملة. على سبيل المثال ، قادة العالم الذين تم تمكينهم بالمراقبة قد يكونون أكثر أو أقل عرضة لاتخاذ قرارات حمقاء تزيد من نقاط الخطورة المحتملة من النوع صفر.

٤٨- وهناك حالة خاصة من نقاط الخطورة المحتملة من النوع ٢ أ هي الحالة التي تعمل فيها بعض الأنظمة بشكل مشترك على تحقيق العتبة المدمرة من خلال إلحاق الضرر بسكانها. لنفترض ، على سبيل المثال ، أن أحدهم كانت لديه نظرة متشائمة للغاية تجاه القادة السياسيين ، ويعتقد أنهم سوف يكونون على استعداد لقتل جزء كبير جدًا من سكانهم إذا كان ذلك سيساعدهم على التمسك بالسلطة أو اكتساب المزيد من المصادر. في حين أن مثل هذه المبادرة للإبادة الجماعية عادة ما تأتي اليوم بنتائج عكسية من وجهة نظر القائد (لأنها من شأنها أن تشعل الثورات وتحطم الاقتصاد) ، يمكن للمرء أن يتخيل بيئة تكنولوجية مختلفة من شأنها أن تخفف هذه القيود – على سبيل المثال ، إذا كان يمكن لقوة شرطة الذكاء الاصطناعي الدولية شديدة المركزية الموثوقة أن تقمع أي مقاومة ، وإذا كان بإمكان الروبوتات أن تحل محل العمال البشر بسهولة. (لحسن الحظ ، يبدو ذلك في كثير السيناريوهات التي تصبح فيها هذه الأشياء مجدية تقنيًا ، فإنها سوف تضعف دوافع الحاكم للقيام بأعمال الإبادة الجماعية: من المفترض أن تمكن قوة شرطة الذكاء الاصطناعي الدولية المفترضة الحاكم من القيام بالحفاظ على السلطة دون قتل أجزاء كبيرة من السكان ، كما أن ميكنة الاقتصاد سوف تزيد الثروة بقدر كبير من أجل أن يكون جزء صغير من الدخل القومي كافيا لتوفير مستوى معيشة مرتفع للمواطنين.)

٤٩ – على سبيل المثال ، غالبا ما تركز مناقشات المراقبة على المفاضلات بين المصالح الخاصة للأفراد ومتطلبات الصالح العام للأمن ضد الهجمات الإرهابية الصغيرة. (حتى الحوادث الإرهابية التي تعتبر عادة كبيرة، مثل هجمات ١١ سبتمبر ، هي على نطاق صغير جدًا وفقًا للمعايير المستخدمة في هذه الورقة.)

٥٠- وفقا لشوارتز Schwartz (١٩٩٨) ، فإن سباق التسلح النووي خلال الحرب الباردة تكلف ٥,٨ تريليون (١٩٩٦ دولار) من النفقات الأمريكية وحدها ، أي ما يعادل ٩,٣ تريليون دولار في ٢٠١٨. هذا التقدير شامل تمامًا ويغطي دورة الوقود ، والأسلحة ، أنظمة التسليم ، ونزع التسليح ، إلخ. إذا أضفنا نفقات البلدان الأخرى ، يمكننا أن نستنتج أن الحضارة الإنسانية أنفقت ما يزيد عن ١٠ تريليون دولار على تطوير وصيانة القدرة على تدمير نفسها بالأسلحة النووية. معظم التكلفة المتكبدة كانت في فترة كان الناتج المحلي الإجمالي العالمي فيها أقل بكثير منه اليوم. حتى أنه تم إنفاق مبالغ أكبر على القدرات العسكرية غير النووية (أكثر من ٢٠ تريليون دولار في ٢٠١٨ في الولايات المتحدة وحدها) . من الممكن أن تكون النفقات النووية قد وفرت المال في الميزانية عن طريق تخفيض الإنفاق العسكري غير النووي. إن النفقات العسكرية النووية وغير النووية تعكس فشل الحضارة الإنسانية في حل مشكلات التنسيق العالمي.

٥١- بشكل جوهري وليس بصورة كلية ، حتى لو تم تفكيك جميع الأسلحة النووية ، فمن المحتمل إنشاء أخرى جديدة.

٥٢ – قد يتطلب الاتفاق على نزع السلاح النووي الكامل ، بطبيعة الحال ، بعض الأحكام المتعلقة بالقوى التقليدية وغير التقليدية كذلك ، حتى لا تعرض الاستقرار الاستراتيجي للخطر.

٥٣- يمكن للمرء أن ينظر إلى أمثلة تاريخية أخرى للحصول على فئة مرجعية أكبر؛ جهود العالم حتى الآن فيما يتعلق بمكافحة الاحتباس الحراري العالمي لا توحي بالثقة في قدرتها على التعامل على وجه السرعة مع مشكلات العمل الجماعي العالمي الأكثر صعوبة. من ناحية أخرى ، جرت معالجة مشكلة استنفاد الأوزون بنجاح من خلال بروتوكول مونتريال.

٥٤. قد يكون الفرض من جانب واحد أسرع ، لكنه يتطلب أن يكون لدى فاعل معين القدرة على فرض إرادته بمفرده على بقية العالم. إذا كان أحد الفاعلين يتمتع بميزة القوة الساحقة ، شكل من أشكال الحكم العالمي بحكم الواقع (الضعيف أو الكامن) القائم بالفعل.

٥٥- على سبيل المثال ، قد يتم تخريب مشروع حسَن النية في إطار تنفيذه ؛ أو قد يتضح أن به أخطاء أو عيوب تصميم تأسيسية لا تظهر إلا بعد فترة من التشغيل الطبيعي. حتى لو كان النظام نفسه يعمل بالضبط على النحو المنشود ويظل غير فاسد ، فقد يلهم مبدع أنظمة المراقبة الأخرى التي لا تتمتع بنفس الضمانات الديمقراطية.

هوامش المترجم

(*) نيك بوستروم Nick Bostrom فيلسوف وعالم متعدد الثقافات مولود في السويد وله خلفية في الفيزياء النظرية ، وعلم الأعصاب الحسابي ، والمنطق ، والذكاء الاصطناعي ، وكذلك الفلسفة. وهو أستاذ في جامعة أكسفورد ، حيث يرأس معهد مستقبل الإنسانية كمدير مؤسس له. وهو مؤلف ما يقرب من ٢٠٠ عمل منشور ، من بينها:

Anthropic Bias (2002) , Global Catastrophic Risks (2008) , Human Enhancement (2009) , and Superintelligence: Paths, Dangers, Strategies (2014)

وهذا الكتاب الأخير من أكثر الكتب مبيعًا في نيويورك تايمز والذي ساعد في إثارة نقاش عالمي حول الذكاء الاصطناعي. لقد ألقى عمل بوستروم المؤثر على نطاق واسع – والذي يجتاز الفلسفة ، والعلوم ، والأخلاق ، والتكنولوجيا – الضوء على الروابط بين أفعالنا الحالية والنتائج العالمية طويلة المدى ، وبالتالي ألقى ضوءًا جديدًا على حالة الإنسان. https://www.nickbostrom.com\

(**) DIY hacking tools = باختصار شديد: حركة اجتماعية للتكنولوجيا الحيوية متنامية ، يدرس فيها الأفراد والمجتمعات والمنظمات الصغيرة علم الأحياء وعلوم الحياة باستخدام نفس الأساليب مثل مؤسسات البحث التقليدية.

(***) عناصر الثقافة ، أو نظام السلوك ، التي يمكن اعتبارها تنتقل من فرد إلى آخر بوسائل غير جينية ، والتقليد بشكل خاص.

(****) تعني مراقبة الكل Panopticon نوعا من السجون قام بتصميمه الفيلسوف الانجليزي والمنظر الاجتماعي جيرمي بنتام Jeremy Bentham عام ١٧٨٥ ، ومفهوم التصميم هو السماح لمراقب واحد بمراقبة جميع السجناء دون أن يكونوا قادرين على معرفة ما إذا كانوا مراقبين أم لا. (المترجم)

(*****) لا أوافق بروفيسور نِك بوستروم في وجهة نظره الخاصة بتحبيذه تحقيق الحوكمة العالمية عن طريق صعود قوة عالمية واحدة إلى مكان الهيمنة التي تمكنها من فرض خطة منفردة من جانبها على بقية العالم؛ (رغم وجاهة الفكرة وقابليتها للتنفيذ) لأن هذا يرجح ويؤيد استبداد هذه القوة بالرأي الذي يخدم مصالحها ويعادي أعداءهما. وما هي القوة الوحيدة عالميا التي لديها من العدالة ما يجعل كل دول العالم تقبل برأيها وتسير في ركابها؟! وأرى أن تعدد المواقف المتحكمة في تكوين وتوجيه وعمل هذه الحوكمة العالمية أكثر ضمانا لترشيد هذه السلطة في تكوينها وتوجيهها وعملها. ربما استخدم بروفيسور بوستروم هذه الفكرة بدافع الرعبة في تسريع اتخاذ القرار بالتدخل لوقف ما تراه مهددا للأمن والسلامة العالميين ، لكن هذا لا يعد مبررا لانفراد دولة واحدة باتخاذ هذا القرار لعدم توفر الحيادية لديها وعدم توفر الموافقة الدولية.

(******) يستخدم بروفيسور بوستروم في هذه المعالجة استعارة من صندوق بندورا الذي خرجت منه الشرور ولا يملك من أخرجها ردها إلى داخل الصندوق مرة أخرى ، وهذه إحدى الاستعارات التي استخدمها باقتدار.

References

- Armstrong, S., Bostrom, N. and Shulman, C. (2016) 'Racing to the Precipice: A Model of Artificial Intelligence Development', AI & Society, 31 (2), pp. 201–06.
- Aviles, L. and Purcell, J. (2012) 'The Evolution of Inbred Social Systems in Spiders and Other Organisms: From Short-term Gains to Longterm Evolutionary Dead Ends?', Advances in the Study of Behavior, 44 (1), pp. 99–133.
- Badash, L. (2001) 'Nuclear Winter: Scientists in the Political Arena', Physics in Perspective, 3 (1), pp. 76–105.
- Banisar, D. (1999) 'Stopping Science: The Case of Cryptography', Health Matrix, 9(2), pp. 253–87.
- Baum, S., de Neufville, R. and Barrett, A. (2018) 'A Model for the Probability of Nuclear War', Global Catastrophic Risk Institute Working Paper 18–1. Available from: https://doi.org/10.2139/ssrn.3137081 [Accessed 31 July 2019].

- Blinder, A. and Perlroth, N. (2018) 'A Cyberattack Hobbles Atlanta, and Security Experts Shudder', New York Times Available from: https://www.nytimes.com/2018/03/27/us/cyberattack-atlanta-ransomware.html
- Bostrom, N. (2002) 'Existential Risks – Analyzing Human Extinction Scenarios and Related Hazards',Journal of Evolution and Technology, 9 (1), pp. 1–30.
- Bostrom, N. (2003) 'Are You Living in a Computer Simulation?', Philosophical Quarterly, 53 (211), pp. 243–55.
- Bostrom, N. (2006) 'What is a Singleton',Linguistic and Philosophical Investigations, 5 (2), pp. 48–54.
- Bostrom, N. (2008) 'Why I Want to be a Posthuman When I Grow Up', in B. Gordijn and R. Chadwick (eds), Medical Enhancement and Posthumanity. New York: Springer, pp. 107–37.
- Bostrom, N. (2011) 'Information Hazards: A Typology of Potential Harms from Knowledge', Review of Contemporary Philosophy, 10(1), pp. 44–79.
- Bostrom, N. (2013) 'Existential Risk Prevention as Global Priority',Global Policy, 4 (1), pp. 15–31.
- Bostrom, N. and _Cirkovi_c, M. M. (eds) (2011) Global Catastrophic Risks. Oxford: Oxford University Press, pp. 1–29.
- Bostrom, N., Douglas, T. and Sandberg, A. (2016) 'The Unilateralists Curse and the Case for a Principle of Conformity', Social Epistemology, 30 (4), pp. 350–71.
- Brooks, S. G. (1999) 'The Globalization of Production and the Changing Benefits of Conquest', Journal of Conflict Resolution, 43 (5), pp. 646–70.
- Brower, M. (1989) 'Targeting Soviet Mobile Missiles: Prospects and Implications', Survival, 31 (5), pp. 433–45.
- Centre for Biosecurity and Biopreparedness (2017) An Efficient and Practical Approach to Biosecurity. Available from: https://www.biosecurity.dk/fileadmin/user_upload/PDF_FILER/Biosecurity_book/An_efficient_and_Practical_approach_to_Biosecurity_web1.pdf [Accessed 31 July 2019].
- Colby, E. A. and Gerson, M. S. (eds.) (2013) Strategic Stability: Contending Interpretations. Carlisle Barracks: U.S. Army War College Press.
- Collingridge, D. (1980) The Social Control of Technology. New York: St. Martin's Press.
- Crafts, N. and Fearon, P. (2010) 'Lessons from the 1930s Great Depression', Oxford Review of Economic Policy, 26 (3), pp. 285–317.

- Danzig, R., Sageman, M., Leighton, T., Hough, L., Yuki, H., Kotani, R. et al. (2011) Aum Shinrikyo: Insights Into How Terrorists Develop Biological and Chemical Weapons, 2nd Edn. Washington, D.C.: Center for a New American Security.
- Day, E. H., Hua, X. and Bromham, L. (2016) 'Is Specialization an Evolutionary Dead End? Testing for Differences in Speciation, Extinction and Trait Transition Rates across Diverse Phylogenies of Specialists and Generalists', Journal of Evolutionary Biology, 29 (6), pp. 1257–67.
- Diamond, J. (2005) Collapse: How Societies Choose to Fail or Succeed. New York: Penguin.
- Drexler, K. E. (1986) Engines of Creation. New York: Anchor.
- Duprex, W. P., Fouchier, R. A., Imperiale, M. J., Lipsitch, M. and Relman, D. A. (2015) 'Gain-of-function Experiments: Time for a Real Debate', Nature Reviews Microbiology, 13 (1), pp. 58–64.
- Ellsberg, D. (2017) The Doomsday Machine: Confessions of a Nuclear War Planner. New York: Bloomsbury Publishing USA.
- Fauci, A. S., Nabel, G. J. and Collins, F. S. (2011) 'A Flu Virus Risk Worth Taking.' The Washington Post, Available from: https://www.washingtonpost.com/opinions/a-flu-virus-risk-worth taking/2011/12/30/gIQAM9sNRP_story.html [Accessed 31 July 2019].
- Federal Select Agent Program ((2017) Suitability Assessment Program Guidance. Available from: https://www.selectagents.gov/resources/Suitability_Guidance.pdf [Accessed 31 July 2019].
- Field, C. B.,Barros, V. R., Mastrandrea, M. D., Mach, K. J., Abdrabo, M. A.-K., Adger, W. N. et al. (2014) Climate Change 2014: Impacts, Adaptation, and Vulnerability. Working Group II Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.
- Ford, C. A. and Rosenberg, D. A. (2005) 'The Naval Intelligence Underpinnings of Reagan's Maritime Strategy', Journal of Strategic Studies, 28 (2), pp. 379–409.
- Franck, J., Hughes, D. J., Nickson, J. J., Rabinowitch, E., Seaborg, G. T., Stearns, J. C. and Szilard, L. (1945) 'Report of the Committee on Political and Social Problems', Manhattan Project 'Metallurgical Laboratory', University of

Chicago, U.S. National Archives, Record Group 77, Records of the Chief of Engineers, Manhattan Engineer District, Harrison-Bundy File, folder #76.
- Friedlingstein, P., Andrew, R. M., Rogelj, J., Peters, G. P., Canadell, J. G., Knutti, R. et al. (2014) 'Persistent Growth of CO2 Emissions and Implications for Reaching Climate Targets', Nature Geoscience, 7 (10), pp. 709-15.
- Gartzke, E. (2007) 'The Capitalist Peace', American Journal of Political Science, 51 (1), pp. 166-91.
- Gartzke, E. and Rohner, D. (2011) 'The Political Economy of Imperialism, Decolonization and Development', British Journal of Political Science, 41 (3), pp. 525-56.
- Greenberg, A. (2012) This Machine Kills Secrets: How Wikileaks, Hacktivists, and Cypherpunks are Freeing the World's Information. New York: Dutton.
- Gronvall, G. K. (2013) H5n1: A Case Study for Dual-Use Research. New York: Council on Foreign Relations.
- Holloway, D. (1994) Stalin and the Bomb: The Soviet Union and Atomic Energy, 1939-1956. New Haven: Yale University Press.
- Igic, B. and Busch, J. W. (2013) 'Is Self-fertilization an Evolutionary Dead End?', New Phytologist, 198 (2), pp. 386-97.
- Inklaar, R., de Jong,H., Bolt, J. and van Zanden, J. L. (2018) 'Rebasing 'Maddison': New Income Comparisons and the Shape of Long-run Economic Development', GGDC Research Memorandum GD-174, Groningen: Groningen Growth and Development Center.
- Jaffe, R. L., Busza, W., Wilczek, F. and Sandweiss, J. (2000) 'Review of Speculative 'Disaster Scenarios' at RHIC', Reviews of Modern Physics, 72 (4), pp. 1125-1140.
- Jones, C. I. (2016) 'Life and Growth', Journal of Political Economy, 124 (2), pp. 539-78.
- Kemp, R. S. (2012) 'SILEX and Proliferation', Bulletin of the Atomic Scientists. Available from: https://thebulletin.org/2012/07/silex-andproliferation/ [Accessed 31 July 2019] .
- Konopinski, E. J., Marvin, C. and Teller, E. (1946) 'Ignition of the Atmosphere with Nuclear Bombs', Report LA-602. Los Alamos, NM: Los Alamos Laboratory.
- Lechner, S. (2017) 'Anarchy in International Relations', Oxford Research Encyclopedia of International Studies [online], Oxford: Oxford University Press. Available from:

https://oxfordre.com/internationalstudies/view/10.1093/acrefore/9780190846626.01.0001/acrefore-9780190846626-e-79?print=pdf [Accessed 31 July 2019].

- Lewis, G., Millet, P., Sandberg, A., Snyder-Beattie, A., and Gronvall, G.(2019) 'Information Hazards in Biotechnology', Risk Analysis, 39(5), pp. 975–81.

- Liberman, P. (1993) 'The Spoils of Conquest', International Security, 18 (2), pp. 125–153.

- Lieber, K. A. and Press, D. G. (2017) 'The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence', International Security, 41 (4), pp. 9–49.

- Lundgren, C. (2013) 'What are the Odds? Assessing the Probability of a Nuclear War', The Nonproliferation Review, 20 (2), pp. 361–74.

- Morland, H. (1979) 'The H-bomb Secret: How We Got It - Why We're Telling It', The Progressive, 43 (11), pp. 3–12.

- Mowatt-Larssen, R. and Allison, G. T. (2010) Al Qaeda Weapons of Mass Destruction Threat: Hype or Reality?. Cambridge, MA: Belfer Center for Science and International Affairs.

- Mueller, J. (2009) Atomic Obsession: Nuclear Alarmism from Hiroshima to al-Qaeda. Oxford: Oxford University Press.

- Munz, P., Hudea, I., Imad, J. and Smith, R. J. (2009) ''When Zombies Attack!: Mathematical Modelling of an Outbreak of Zombie Infection', in J.M., Tchuenche and C., Chiyaka (eds.), Infectious Disease Modelling, Research Progress. New York: Nova Science Publishers, pp. 133–50.

- Norris, R. S. and Kristensen, H. M. (2010) 'Global Nuclear Weapons Inventories, 1945–2010', Bulletin of the Atomic Scientists, 66 (4), pp. 77–83.

- Olson, K. B. (1999) 'Aum Shinrikyo: Once and Future Threat?', Emerging Infectious Diseases, 5 (4), pp. 413–16.

- Ord, T., Hillerbrand, R. and Sandberg, A. (2010) 'Probing the Improbable: Methodological Challenges for Risks with Low Probabilities and High Stakes', Journal of Risk Research, 13 (2), pp. 191–205. Global Policy (2019) 10:4 © 2019 The Authors

- Parfit, D. (1987) Reasons and Persons., Oxford: Clarendon Press.

- Parker, E. R. and Jacobs, L. G. (2003) 'Government Controls of Information and Scientific Inquiry', Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science, 1 (2), pp. 83–95.

- Persson, I. and Savulescu, J. (2012) Unfit for the Future: The Need for Moral Enhancement. Oxford: Oxford University Press.
- Re, R. M. (2016) 'Imagining Perfect Surveillance.' UCLA Law Review Discourse, 64(1), pp. 264–92.
- Rhodes, R. (1986) The Making of the Atomic Bomb. London: Simon and Schuster.
- Sagan, S. D. (1995) The Limits of Safety: Organizations, Accidents, and Nuclear Weapons. Princeton, NJ: Princeton University Press.
- Sagan, S. D. and Waltz, K. N. (2012) The Spread of Nuclear Weapons: An Enduring Debate. New York: WW Norton.
- Sanchez, J. (2013) 'A Reply to Epstein and Pilon on NSA's Metadata Program', Cato at Liberty. Available from: https://www.cato.org/blog/reply-epstein-pilon-nsas-metadata-program [Accessed 31 July 2019].
- Schelling, T. C. (1960) The Strategy of Conflict. Cambridge, MA: Harvard University Press.
- Schlosser, E. (2013) Command and Control: Nuclear Weapons, the Damascus Accident, and the Illusion of Safety. New York: Penguin Press.
- Schwartz, S. I. (1998) Atomic Audit: The Costs and Consequences of US Nuclear Weapons Since 1940. Washington, D.C.: Brookings Institution Press.
- Sharp, P. A. (2005) '1918 Flu and Responsible Science', Science, 310 (5745), p. 17.
- Shindell, D. T. (2014) 'Inhomogeneous Forcing and Transient Climate Sensitivity', Nature Climate Change, 4 (4), pp. 274–77.
- Sokoski, H. D. and Tertrais, B. (2013) Nuclear Weapons Security Crisis: What Does History Teach?. Washington, D.C.: Nonproliferation Policy Education Center.
- Stevenson, J. (2008) Thinking Beyond the Unthinkable: Harnessing Doom from the Cold War to the War on Terror. New York: Viking.
- Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M. M. B., Allen, S. K., Boschung, J. et al. (2014) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. New York: Cambridge University Press.
- Swire, P. (2015) The Declining Half-Life of Secrets and the Future of Signals Intelligence. Washington, D.C.: New America.
- Tegmark, M. and Bostrom, N. (2005) 'Astrophysics: Is a Doomsday Catastrophe Likely?', Nature, 438 (7069), p. 754.

- Torres, P. (2018) ‘Superintelligence and the Future of Governance: On Prioritizing the Control Problem at the End of History.’ in R. Yampolskiy (ed.), Artificial Intelligence Safety and Security. Milton: Chapman and Hall/CRC, pp. 357–74.

Author Information

Nick Bostrom is a Professor at Oxford University, where he directs the Future of Humanity Institute, which includes the Center for the Governance of AI and research groups working on other macrostrategic challenges for humanity.