مشاركة العالم مع العقول الرقمية (1) (*)

نِك بوستروم ، كارل شولمان (**)

ترجمة دكتور/ حسين عبدالغني إبراهيم

نبذة مختصرة **Abstract**

تشغل عقول المخلوقات البيولوجية ركنًا صغيرًا من مساحة أكبر بكثير يمكن أن تشغلها العقول المحتملة التي يمكن إنشاؤها بمجرد إتقان تقنية الذكاء الاصطناعي. ومع ذلك ، فإن الكثير من حدسنا وممارساتنا الأخلاقية تستند إلى افتراضات حول الطبيعة البشرية التي لا نحتاج إلى الاحتفاظ بها في العقول الرقمية. هذا يشير إلى الحاجة إلى التفكير الأخلاقي مع اقترابنا من عصر الذكاء الآلي المتقدم. وهنا ، نركز على مجموعة من القضايا ، والتي تنشأ من منظور "وحوش المنفعة" الرقمية. قد يجري إنتاج هذه العقول بكميات كبيرة وذات أوضاع واهتمامات أخلاقية مماثلة لتلك الخاصة بالبشر أو الحيوانات الأخرى ذات الأهمية الأخلاقية ، بحيث تفوق متطلباتها الأخلاقية بشكل جماعي متطلبات البشر الحاليين. أضف إلى ذلك أنه قد يصبح من السهل إنشاء عقول رقمية فردية ذات اهتمامات فردية ومتطلبات من الموارد أكبر بكثير من متطلبات البشر. إن عدم تقدير هذه الاحتمالية يمكن أن ينتج عنه كارثة أخلاقية ذات أبعاد هائلة، في حين أن الطريقة الساذجة في تقديرها يمكن أن تكون كارثية على الإنسانية. إن النهج المعقول يتطلب إصلاح معاييرنا الأخلاقية ومؤسساتنا جنبًا إلى جنب مع التخطيط المسبق الخاص بأنواع العقول الرقمية التي ننشؤها.

1- مقدمة Introduction

تفرض الطبيعة البيولوجية البشرية الكثير من المحددات العملية لما يمكن فعله لتعزيز رفاهية شخص معين. لا يرضينا سوى أن نعيش طويلاً ، ونشعر بفرح كبير ، وننجب الكثير من الأطفال ، ونستفيد كثيرًا من الدعم والموارد الإضافية. في غضون ذلك ، من أجل الازدهار ، نحتاج إلى تلبية مجموعة معقدة من الظروف الجسدية ، والنفسية ، والاجتماعية.

في المقابل ، يمكننا بسهولة تصور كائنات اصطناعية ذات تجارب واعية ، ورغبات ، وقدرة على التفكير واتخاذ القرارات المستقلة ، والتي لن تخضع لـ ذات المحددات العملية في قدرتها على الاستفادة من الموارد الإضافية ولن تعتمد على ذات المتطلبات المعقدة من أجل بقائها وازدهارها. قد يكون هذا تطورًا رائعًا: حياة خالية من الألم والمرض ، بل مليئة بالسعادة ، وغنية بالوعي والفهم الخارقيْن وكل أنواع المنتجات الراقية. (2)

يثير التقدم الأخير في دراسة الآلات احتمال أن تتمتع أنظمة الذكاء الاصطناعي المستقبلية قريبًا (أو بالفعل) بالقدرات النفسية الموجودة عادةً في البشر أو الحيوانات الأخرى لمنح تلك الأنظمة درجات من المكانة الأخلاقية. (سوف نفترض أن الذكاء الاصطناعي المصمم بشكل مناسب يمكن أن يكون واعيًا ، على الرغم من أنه ومن الجدير بالذكر أن البعض لا يعيرون اهتماما للحالة الأخلاقية ولا يرون التمتع بالوضع الأخلاقي شرطًا ضروريًا). (3)

هذا يمهد الطريق أمام كتاب روبرت نوزيك Robert Nozick (1974 ، ص 41) "وحش المنفعة Utility Monster" ليصبح حقيقة عملية:

تشعر النظرية النفعية بالحرج من احتمالية حصول وحوش المنفعة على قدر أكبر بكثير من المنفعة من التضحيات التي يبذلها الآخرون أكثر مما يخسره الآخرون. لأنه ، بشكل غير مقبول ، يبدو أن هذه النظرية تتطلب أن تتم التضحية بنا جميعًا في فم الوحش ، من أجل زيادة المنفعة الكلية.

لقد كان الاتساق المفاهيمي لوحوش المنفعة محل نزاع. على سبيل المثال ، كتب ديريك بارفيت Derek Parfit (1984 ، ص 343):

إذا فكرنا في حياة الإنسان الأكثر حظًا بين الذين نعرفهم ، وسألناه كيف يجب أن تكون الحياة حتى تستحق أن يعيشها الإنسان مليون مرة، نرى أن الفجوة النوعية بين تلك الحياة وحياتنا ، في أفضل حالاتها ، لابد أن تشبه الفجوة بين حياتنا ، في أفضل حالاتها ، وحياة تلك المخلوقات الواعية بالكاد – مثل ، "الصامتون القانعون contented oysters" لأفلاطون ، إذا كانوا واعين. يبدو أن القول بأننا لا يمكننا تخيل حياة وحش المنفعة هذا ، حتى في أضعف صوَرِها ، يعد ردًّا مقبولا. وهذا يلقي بظلال من الشك على قوة المثال. ربما يقول علماء المنفعة أننا ، إذا استطعنا

حقًا تخيل ما ستكون عليه مثل هذه الحياة ، قد لا نجد اعتراض نوزيك مقنِعًا. حيث يبدو أن "الوحش" الذي يتخيله يعد كائنًا يشبه الإله. وفي ظل الوجود المتخيل لمثل هذا الكائن ، فإن إيماننا بحقنا في المساواة معه قد يبدأ بالتردد ...

ومع ذلك ، على الرغم من أن بارفيت Parfit يشكك في إمكانية حدوث اختلاف نوعي هائل في الرفاهية ، فإنه يقترح أنه يمكن الحصول على نتائج مماثلة من خلال النظر في البعد الكمي لحجم السكان ، حيث من الواضح أنه لا يوجد حاجز مفاهيمي للقيم القصوى.

سوف نقول بأن حجم السكان هو واحد فقط من عدة أبعاد كمية – جنبًا إلى جنب مع العديد من الأبعاد النوعية الأقل تحديدًا – والتي من خلالها يمكن لوحوش المنفعة أن تتفوق على البشر والحيوانات الأخرى من خلال السماح بقدر أكبر من الرفاهية في كل وحدة من استهلاك الموارد. هذا الوجود متعدد المسارات المستقلة لدى وحوش المنفعة يجعل الاستنتاج أكثر مصداقية بأن وحوش المنفعة هذه ممكنة وقد تصبح ممكنة عمليًا مع مزيد من التقدم في الذكاء الاصطناعي.

وفي حين أن غير النفعيين قد يتوهمون أنفسهم محصنين ضد تحدي وحش المنفعة ، فإن معظم الآراء المعقولة هي في الواقع عرضة للتأثر بدرجات مختلفة. هذا لأنه حتى لو افترضنا أنه لن تحدث انتهاكات أخلاقية (مثل القتل المباشر) ، فقد تكون المصالح البشرية لاتزال متأثرة سلبًا بظهور وحوش المنفعة ، حيث يمكن أن توجد للأخيرة مطالبات أخلاقية أقوى لمساعدة الدولة والموارد النادرة الأخرى ، وبالتالي تقليل المقدار الذي يمكن للبشر المطالبة به بشكل دفاعي.

لذا فإن الشيء الجيد في وحوش المنفعة هو أنها يمكن أن تولِّد الكثير من الفوائد. والشيء السيئ هو أننا قد نتحمل عواقب هذا الأمر.

2- بنية وحوش المنفعة The anatomy of utility monsters

2-1- القدرة على إعادة إنتاج ذاتها Reproductive capacity

إحدى الميزات الأساسية لبرامج الكمبيوتر تتمثل في سهولة وسرعة الاستنساخ الدقيق ، بشرط توفر أجهزة الكمبيوتر. يمكن بناء الأجهزة بسرعة طالما أن ناتجها الاقتصادي يمكن أن يدفع تكاليف التصنيع (التي انخفضت تاريخيًا ، على أساس السعر والأداء ، بمبالغ هائلة ؛ (نوردهاوس Nordhaus 2007) . هذا يفتح الباب أمام ديناميكيات السكان التي قد تستغرق عدة قرون لكي تتحول بين البشر ويتم ضغطها في جزء صغير من عمر الإنسان. حتى لو كان من الممكن في البداية بناء عدد قليل من العقول الرقمية ذات قدرة فكرية معينة بتكلفة معقولة ، فإن عدد هذه العقول يمكن أن ينمو بسرعة بشكل متضاعف أو أضعافا مضاعفة ، حتى يتم تقييدها بواسطة محددات أخرى. إمكانية الاستنساخ المتفجرة هذه يمكن أن تسمح للعقول الرقمية بأن تفوق عدد البشر بشكل كبير في وقت قصير نسبيًا – مما يؤدي بالمقابل إلى زيادة القوة الجماعية الخاصة بمطالبها الأخلاقية والسياسية.

علاوة على ذلك ، إذا استمر إنتاج العقول الرقمية والأجهزة المطلوبة حتى تتساوى أجور العقول الناتجة مع التكاليف الهامشية ، فقد يؤدي ذلك إلى انخفاض الأجور نحو مستويات كفاف الآلة حيث تصبح الموارد الطبيعية عاملاً مقيِّدا. وقد تكون غير كافية للبشر للبقاء على قيد الحياة (والعقول الرقمية القديمة) (Hanson,2001 ; Aghion, Jones and Jones, 2017 . مثل هذه الظروف تجعل قضايا إعادة التوزيع أكثر إلحاحًا – مسألة حياة أو موت – في حين أن النمو السكاني المالتوسي (نسبة إلى الاقتصادي ورجل الدين الإنجليزي توماس روبرت مالتوس) قد يجعل المطالبات بتحويل المدفوعات نهمًا فعليًا.

جانب آخر مزعج للاستنساخ السريع والرخيص هو أنه يسمح بتغيير سريع للسكان. فيمكن أن تحل نسخة من عقل كامل من أحدث إصدار محل العقل الرقمي الذي تم حذفه على الفور – بعكس الحالة البشرية ، (حيث يستغرق إنتاج طفل يسيل لعابه تسعة أشهر). (4) ويمكن للضغوط الاقتصادية بالتالي الدفع باتجاه محو متكرر جدًا للعقول "القديمة" وإحلال العقول التي تولد قيمة اقتصادية أكبر بنفس البنية محلها. (5)

وبالتالي ، فإن استمرارًا معقولًا لممارسات البرمجيات الحالية المطبقة على العقول الرقمية يمكن أن ينطوي على أعداد كبيرة جدًا من الأعمار القصيرة والوفيات ، حتى كجزء من عدد العقول الموجودة في أي وقت. قد تكون هذه العقول الرقمية سريعة الزوال ناضجة نفسياً ، وشابة زمنياً ، ولديها عمر طويل محتمل ، لكن متوسط العمر الافتراضي يكون قصير للغاية في غياب الدعم. إذا كنا نعتقد أن احتضار الشباب رغم إمكانية العيش لفترة طويلة يمثل حرمانًا كبيرًا ، أو أنه غير عادل للغاية بينما يتمكن الآخرون من عيش حياة طويلة ، فقد يؤدي هذا إلى مطالبة هذه العقول الرقمية بالموارد لإطالة عمرها (أو غيرها من أشكال التعويضات). إذا كان الموت في حد ذاته أمرًا سيئًا (وليس مجرد تكلفة لفرصة للحياة الضائعة) ، فإن هذا التغير السريع في العقول يمكنه أن يزيد أيضًا من مدى هذا الاستخفاف بكل عام تعيشه من الحياة.

2-2- تكلفة المعيشة Cost of living

من المعقول أن يحتاج الكثير من العقول الرقمية إلى القليل من الدخل للحفاظ على مستوى معيشي معين. فمن المرجح أن تنخفض تكلفة أجهزة الكمبيوتر لدعم العقول الرقمية إلى ما دون تكلفة دعم العقل والجسم البشري. إذا نظرنا إلى ما هو أبعد من مجرد الكفاف ، والسلع المادية والخدمات المناسبة لاستهلاك الإنسان (مثل السكن والنقل) فمن المحتمل أن تكون أكثر تكلفة من تكنولوجيا المعلومات والسلع الافتراضية لتلبية الاحتياجات المكافئة بالنسبة للعقل الرقمي. فنحن لا تحتاج إلى عقل رقمي يعاني من ظروف بيئية قاسية أو تلوث أو مرض أو شيخوخة بيولوجية أو أي عدد من الإكراهات الأخرى التي تقلل من رفاهية الإنسان.

من المحتمل أن تكون تكلفة إنتاج عدد معين من سنوات الحياة (معدلة الجودة) لعقل رقمي شبيه بعقل الإنسان أقل بكثير من التكلفة المكافئة بالنسبة لعقل الإنسان البيولوجي. إن الفروق الكبيرة في تكلفة المعيشة تعني أن ، عند ظهور مسائل التوزيع ، المورد الذي يحقق فائدة صغيرة للإنسان قد يحقق فوائد كبيرة للكثير من العقول الرقمية. وإذا كانت ميزانية الطاقة المطلوبة للحفاظ على حياة بشرية واحدة لمدة شهر يمكن أن تحافظ على عشرة عقول رقمية لمدة عام ، فإن ذلك من شأنه أن يقيم حجة قوية لتفضيل الأخيرة في حالة الندرة.

2-3- السرعة الذاتية Subjective speed

يمكن استخدام الأجهزة ذات السرعات التسلسلية العالية لتشغيل العقول الرقمية بشكل أسرع. تُقاس سرعات ساعة الكمبيوتر الحالية بالجيجاهيرتز gigahertz ، وهي أكبر بملايين المرات من معدلات إطلاق الخلايا العصبية البشرية شحناتها ؛ ويمكن أن تتجاوز سرعات نقل الإشارة بالمثل سرعة توصيل الأعصاب البشرية. ولذلك فمن المرجح أن العقول الرقمية ذات القدرات الشبيهة بالبشر يمكنها أن تفكر ، على الأقل آلاف المرات (وربما الملايين) ، أسرع مما يفكر البشر ، وهذا في حالة إمدادها بما يكفي من الأجهزة. إذا قام العقل الرقمي بحزم آلاف السنين من الحياة الذاتية في سنة تقويمية واحدة ، عندها يبدو أن الأول (الوقت الذاتي ، وليس وقت ساعة الحائط) هو المقياس الصحيح لأشياء مثل مقدار الرفاهية المكتسبة من الحياة الممتدة. (Bostrom and Yudkowsky, 2014) .

ونظرًا لأن التسريع يتطلب دفع ثمن المزيد من الأجهزة ، فإن هذا يوفر طريقة للعقول الرقمية الفردية للحصول على عوائد أعلى بكثير (سنوات حياة ذاتية لكل دولار) من الثروة أكثر مما يستطيع البشر تحقيقه عادةً. وعند السرعات المنخفضة ، ستكون المكاسب المتاحة للعقول الرقمية قريبة من المطردة ؛ على الرغم من اقتراب السرعات من حدود التكنولوجيا ، فإن التكاليف الهامشية لزيادة السرعة سوف ترتفع. (6)

ونظرًا لأن هذه المكاسب من السعي بشكل أسرع يمكن أن تعود إلى الأفراد الموجودين في ذلك الوقت والذين كانوا يعملون بشكل أبطأ في البداية ، فإن هذا البعد لدى وحش المنفعة ينطبق أيضًا على البديهيات السكانية التي تتخذ نهج "التأثير على الفرد" (سوف نأتي إلى المزيد حول هذه النقطة لاحقًا) .

2-4- منحنى اللذة Hedonic skew

تعكس الأبعاد المذكورة أعلاه الخصائص الأساسية للعقول المطبقة في أجهزة الكمبيوتر الرقمية ؛ ومن ثم فإنها تنطبق حتى على عمليات المحاكاة عالية الدقة للأدمغة البشرية غير المعدلة (في جسم افتراضي وبيئة افتراضية) . فيتم فتح أبعاد إضافية للرفاهية المتزايدة إذا نظرنا في إمكانية تعديل "برنامج software" ، أو بنية العقل.

يعتمد تحديد هذه التحسينات المتعلقة بالبرمجيات ذات الصلة على نظرية الرفاهية التي نتبناها. في القسم والذي يليه سوف ننظر في مفهوم متعة الرفاهية البسيط. وفي وقت لاحق ، نناقش بعض المفاهيم الأخرى.

إن أدمغتنا ليست مصممة بغرض الرفاهية. على أية حال ، لقد تطور علم وظائف الأعضاء البشري لكي يولد المتعة والألم حيث كانت هذه السلوكيات المحفزة مرتبطة باللياقة الإنجابية في الأجيال الماضية. وهذا يجلب لنا قدرًا كبيرًا من المعاناة التي يصعب تجنبها. علاوة على ذلك ، تميل أوجه متعتنا إلى التنظيم المحكم وإلى الاستغناء عن التحفظ البائس. ملذات الطهي ينظمها الجوع ؛ والملذات الجنسية محدودة بالرغبة الجنسية. والتوفير الشامل لأوجه المتعة المرتبطة بالرغبات بسبب القوة النسبية والوضع الاجتماعي يعوقه الطابع شبه الصفري للسلع المعنية. ويتم التحكم في معظم المكافآت أيضًا من خلال آليات مثل الملل وفتور الرغبة ، والتي تقلل تدريجياً من البهجة الناتجة عن المحفزات المتكررة أو الظروف الجيدة المستمرة.

يمكن تصميم العقول الرقمية لتجربة حالات عقلية ممتعة متواترة ولمدة أكبر ، ولديها ذروة تجارب لمعظم الوقت أو طوال الوقت. كما يمكنها الاستمتاع بالرفاهية الدائمة وكذلك التحرر من الأجزاء المؤلمة في وجودنا البشري الحالي.

سوف يكون التوازن في الملذات لدى البشر ، أيضًا ، قابلاً للتحسين الكبير في نوع التكنولوجيا المتقدمة التي من المحتمل أن تسبق أو تتبع تقنية ذكاء الآلة الناضجة. (7) ومع ذلك ، فإن الضبط الجذري لتوازن المتعة لدى البشر البيولوجيين قد يكون أكثر "تكلفة" من فعل الشيء نفسه بالنسبة للعقول الرقمية الجديدة ، من خلال طريقتين: (أ) التدخلات التي تتطلب جراحة الدماغ ، والضبط والتعديل الدوائي الدقيق ، أو ما يعادله ، وهذا ربما ، على الأقل في المدى القريب ، يكون غير عملي أو مكلف ؛ و (ب) قد تؤدي التحولات الجذرية لطبيعتنا الجسدية إلى المخاطرة بتدمير الهوية الشخصية أو الخصائص الأخرى لطبيعتنا البشرية الحالية التي نقدرها عالياً. (8) وبالتالي يمكن أن يكون للتصميمات الذهنية للآلات الواعية مزايا كبيرة من حيث الكفاءة التي يمكنها باسطتها إدراك الحالات الممتعة ذات القيمة.

2-5- نطاق المتعة Hedonic range

بالإضافة إلى تغيير جزء من الوقت الذي يقضيه المرء في العيش في أجزاء مختلفة من مقياس المتعة المتاحة للبشر الحاليين ، فقد يكون من الممكن أيضًا – بشكل أكثر تخمينًا – تصميم عقول رقمية يمكنها إدراك حالات من مستويات رفاهية المتعة الجيدة "خارج الرسوم البيانية" والتي لا تستطيع أدمغة الإنسان تمثُّلها.

تعطي الاعتبارات التطورية بعض الدعم لهذه الفرضية؛ بقدر ما تتوافق شدة الملذات والآلام مع قوة الاستجابات السلوكية ، ينبغي أن يميل التطور إلى ضبط تجارب المتعة للحصول على درجات جهد معظمة للياقة البدنية تقريبًا لتحقيقها أو تجنبها. ولكن بالنسبة للبشر ، من الأسهل ، بشكل عام ، خسارة قدر كبير من اللياقة الإنجابية في وقت قصير أكثر من اكتساب نفس المقدار. إن البقاء في حريق لبضع لحظات يمكن أن يؤدي إلى إصابة دائمة أو وفاة ، بما يؤثر على جميع فرص الإنجاب المتبقية للكائن الحي. لا توجد وجبة واحدة أو فعل جنسي واحد لديه نفس الإمكانيه في حالة خضوعه للقياس في الثانية – يستغرق الأمر أسابيع لتجويعه أو القضاء عليه ، كما أن العدد المتوقع لإنتاج الأطفال الذين يجري إنتاجهم في دقيقة واحدة من التزاوج يعد صغيرا. وبالتالي ، ربما كان للتطور أن يولد آلامًا محفزة بشكل مكثف في الثانية استجابةً للإصابة أكثر من الملذات عند الاستجابة للأحداث الإيجابية. وعلى النقيض من ذلك ، يمكن صياغة العقول المهندسة لتجربة الملذات ذات المكافأة الشديدة بنفس مقدار أسوأ أنواع العذاب الذي لا يبعث على المكافأة. وقد تكون الرفاهية أو البؤس الواقع خارج التجربة الإنسانية كليةً أمرًا ممكنًا أيضًا.

2-6- تفضيلات غير مكلفة Inexpensive preferences

بالنسبة لتحليلات متعة الرفاهية ، لاحظنا إمكانية صنع وحوش المنفعة من خلال تصميم عقول رقمية إما للعثور على المزيد من الأشياء الممتعة أو للحصول على ملذات قوية فوق طاقة البشر. وبالنسبة لتحليلات التفضيل والرضا بخصوص الرفاهية ، ينشأ ذوج موازٍ من احتمالات نشأة الوحش: تكوين عقول رقمية لها تفضيلات يسهل إرضائها ، أو تكوين عقول رقمية لديها

تفضيلات قوية فوق طاقة البشر. نناقش هنا العقول ذات التفضيلات التي يمكن إرضاؤها بسهولة ، ونرجئ مناقشة الاحتمال الأخير إلى القسم التالي.

الحالة الأساسية واضحة ومباشرة – أكثر من الحالة الموازية المتعلقة بالتجارب الممتعة ، لأن إسناد التفضيلات لا يتطلب افتراضات مثيرة للجدل حول وعي الآلة. إذا فهمنا التفضيلات بطريقة وظيفية ، ككيانات مجردة تشارك في تفسيرات مُرضية لـ (جوانب) سلوك العمليات الذكية الموجهة نحو الهدف (جنبًا إلى جنب مع المعتقدات) ، سوف يكون من الواضح أن العقول الرقمية يمكن أن يكون لها تفضيلاتها. علاوة على ذلك ، يمكن تصميمها مزودة بتفضيلات بحيث يكون من السهل إشباعها بسهولة: على سبيل المثال ، تفضيل وجود إشارة مميّزة على الأقل ، أو الضغط على زر أحمر معين مرة واحدة على الأقل.

تفرض بعض تحليلات إرضاء التفضيلات وجود متطلبات إضافية يمكن أن تعتمد عليها تفضيلات رفاهية شخص ما. على سبيل المثال ، غالبًا ما يتم استبعاد التفضيلات السادية أو الحاقدة. أيضا ، قد يستبعد بعض الفلاسفة التفضيلات "غير المعقولة" ، مثل تفضيل لدى شخص ملتزم باهتمام شديد بحساب كل أوراق العشب في مروج برينستون . (10) اعتمادًا على مدى التزام الفرد بشأن التفضيلات التي يُسمح بها باعتبارها "معقولة" ، وقد يكون هذا من السهل إجراؤه أو قد لا يكون كذلك.

بعض الأنواع الأخرى من المتطلبات التي ربما تفرض أن التفضيلات التي تساهم في الرفاهية يجب أن يتم إقرارها بشكل شخصي (ربما عن طريق جعلها مصحوبة بتفضيل من الدرجة الثانية بهدف الوصول إلى تفضيل من الدرجة الأولى) أو تستند إلى سمات نفسية أو سلوكية إضافية (مثل الميل للابتسام ، والشعور بالتوتر ، وتجربة الفرح ، وأن تصبح مسبوقا من أقرانك ، وجذب انتباه المرء ، وما إلى ذلك). ومع ذلك ، من المحتمل أن يلبي العقل الرقمي هذه المتطلبات. إن البشر لديهم تفضيلات للمتع الحسية ، والحب ، والمعرفة ، والتواصل الاجتماعي، والإنجاز ، والتي يُعتقد أن إرضاءها يساهم في الرفاهية. نظرًا لأنه يمكن بسهولة إنشاء النظائر القريبة من هذه الأشياء في الواقع الافتراضي ، إلى جانب أي سمات نفسية أو سلوكية والتفضيلات من الدرجة الثانية التي قد تكون مطلوبة ، فمن غير المرجح أن تمنع هذه المتطلبات تكوين كائنات ذات تفضيلات قوية ومؤهلة لأن يمكن إرضاؤها بسهولة.

2-7- قوة التفضيل Preference strength

في حين يعد إنشاء تفضيلات سهلة الإرضاء للغاية أمرا بسيطا من الناحية المفاهيمية ، فإن إنشاء تفضيلات ذات "قوة" خارقة يمثل مشكلة أكبر. في تكوين معيار فون نيومان – مورجنسترن Von Neumann–Morgenstern ، تعد وظائف المنفعة فريدة من نوعها فقط بسبب التحويلات الموازية: إضافة إلى دور المنفعة أو مضاعفتها بواسطة عامل ثابت لا يؤثر على الخيارات ، وقوة التفضيل يتم تحديدها فقط بما يتعلق بالتفضيلات الأخرى لنفس الفرد. وبالتالي ، لإجراء مقارنات بين الأفراد ، يجب توفير بعض الهياكل الإضافية لتطبيع وظائف المنفعة المختلفة ووضعها في نطاق مشترك. (11)

هناك الكثير من الأساليب التي تحاول إعطاء "كلمة متساوية" لتفضيلات أشخاص مختلفين بناءً على تكوين التفضيل فقط ، بما يعادل التأثير المتوقع للعوامل المختلفة واستبعاد وحوش المنفعة التي لها قوة التفضيل. (12) هذه الأساليب التي تتغاضي عن بعض الاعتبارات:

أولاً ، لا تأخذ في الاعتبار التعقيد النفسي أو الكفاءات: قد يكون لبعض الأنظمة البسيطة ، مثل منظم الحرارة الرقمي ، نفس وزن العقول المعقدة نفسياً.

ثانيًا ، تنكر أي دور للمعاني العاطفية أو السمات الأخرى التي نستخدمها حدسيًا لتقييم قوة الرغبة في أنفسنا وفي البشر الآخرين.

وثالثًا ، يمكن أن تفشل وظيفة الرفاهية الاجتماعية الناتجة في توفير أساس مقبول للطرفين لتعاون الأطراف غير المهتمة ، لأنها تزود أشخاصا أقوياء ببدائل قوية لها نفس الوزن مثل تلك التي ليس لديها قوة وبدائل.

قد تتطلب المسألتان الأوليان إجراء بحث في ميزات وزن القوة النفسية هذه. وربما يجري التعامل مع المسألة الثالثة من موقع تعاقدى يحدد الأوزان بناءً على الاعتبارات النظرية والمساومة (الافتراضية) للعملية. لن تهيمن وحوش النفعة على النهج التعاقدي بما لا يتناسب مع قوتها التفاوضية ، ولكنها سوف تقترب بشكل خطير من "احتمالية صنع الصواب" ، وسوف تفشل في توفير التوجيه للأطراف المتعاقدة التي تهتم برفاهية العقول الضعيفة وتتمنى مساعدتها.

2-8- سلع القائمة الموضوعية وازدهارها Objective list goods and flourishing

تزعم نظريات القائمة الموضوعية للرفاهية أن مدى جودة حياة شخص ما بالنسبة لها يعتمد على الدرجة التي تحتوي بها حياته على أنواع مختلفة من السلع (والتي قد تشمل ، على سبيل المثال لا الحصر ، المتعة ورضا التفضيل). بعض السلع التي تظهر بطريقة شائعة هي المعرفة ، والإنجاز ، والصداقة ، والفضيلة الأخلاقية ، والتقدير الجمالي ، على الرغم من وجود تباين كبير في تحديد وترجيح السلع المختلفة. يمكن لهذه السلع أن تكون حقائق "موضوعية" في ما يخص كيفية سير حياة شخص ما ، أي أن رفاهية الفرد ليست محددة حصريًا من خلال ما يشعر به أو ما يفضله ولكن أيضًا من خلال قابلية حياته لأن تلبي معايير النجاح الخارجية المختلفة.

يبدو أن الكثير من العناصر التي يمكن العثور عليها في القوائم الموضوعية مفتوحة للتطبيقات المتطرفة في العقول الرقمية. على سبيل المثال ، يمكن تجسيد الفضائل الفكرية في الآلات فائقة الذكاء بدرجات أعلى بكثير مما هي عليه في البشر. الفضائل الأخلاقية ، أيضًا ، ستكون شيئًا يمكن للعقول الرقمية أن تتصف به إلى درجات فوق طاقة البشر: فقد يكون من الممكن إنشاؤها بحيث تبدأ بالمعرفة الأخلاقية الكافية والدافع المثالي دائمًا لفعل ما هو صحيح أخلاقيًا ، حتى تظل معصومة من العيوب ، في حين ينتهي كل إنسان بالغ بسجل أخطاء من المخالفات.

الصداقة سلعة معقدة ، ولكن ربما يمكن اختزالها إلى مكوناتها الأساسية ، مثل الولاء ، والتفاهم المتبادل لشخصيات ومصالح الآخرين بعضهم البعض ، وتاريخ التفاعل الماضي. يمكن عندئذٍ إعادة تجميع هذه المكونات في شكل فعال للغاية ، بحيث يمكن للعقول الرقمية أن تحافظ على أعداد أكبر بكثير من الصداقات الأعمق بدرجة أكبر وعلى مدى فترات أطول كثيرا مما يستطيع البشر القيام به.

أو إذا فكرنا في الإنجاز، وفقًا لتفسير هوركا و تازيولاس Hurka and Tasioulas للإنجاز في كتابهما عام 2006 ، فإن قيمته تعكس الدرجة التي ينتج عنها من ممارسة العقل العملي: أفضل الإنجازات هي تلك التي يتم فيها تلبية الأهداف الصعبة عبر خطط هرمية تنقسم

إلى خطط فرعية أكثر تعقيدًا .. يمكننا عندئذٍ بسهولة تصور "وحوش الإنجاز" الرقمية التي تتابع بلا هوادة مشاريع أكثر تفصيلاً دون أن تكون مقيدة بإيقاف التحفيز أو جذب الانتباه.

بهذه الطرق والعديد من الطرق الأخرى ، يمكن للعقول الرقمية أن تحقق مجموعة متنوعة من السلع الموضوعية إلى مدى أكبر بكثير مما هو في استطاعتنا نحن البشر.

وجهة نظر أخرى للرفاهية هي أنها تتكون من "الازدهار" ، والذي قد يمكن الوصول إليه من خلال ممارسة قدراتنا المميزة أو من خلال تحقيق "غاياتنا". بمفهوم أرسطي ، على سبيل المثال ، يزدهر الكائن إلى الدرجة التي ينجح بها في تحقيق هدفه أو طبيعته الأساسية. قد يبدو هذا النوع من الازدهار متاحًا للعقل الرقمي ، الذي يمكنه بالتأكيد ممارسة قدرات مميزة ، والتي يمكن أيضًا أن تُنسب إلى غاية بمعنى ما لدى البشر – إما غاية محددة بواسطة نوايا المبدع ، أو مشتقة من الديناميكيات التطورية أو الديناميكيات الأخرى التي أوجدتها وشكلت طبيعتها. لذلك ينبغي أن تكون ممكنة على الأقل مساوية ، وربما تتجاوز البشر ، إلى حد ما ، من حيث تحقيق هذا الازدهار ؛ على الرغم من أن الكيفية التي نفهم بها ازدهار الإنسان الخارق بشكل جذري ، طبقا لهذا النوع من التحليل ، تعد أقل وضوحًا.

2-9- مقياس العقل Mind scale

على المستوى المجرد ، يمكننا النظر في مجموعة من المقاييس الذهنية الممكنة ، من العقول الصغيرة الشبيهة بالحشرات (أو حتى التي تشبه الثرموستات) إلى العقول فائقة الذكاء ذات الإنتاجية الحسابية التي تفوق مجموع قدرات البشر اليوم. وتزداد تكلفة تشييد هذه العقول كلما صعدنا في هذا المقياس ، وكذلك تزداد الأهمية الأخلاقية. السؤال المهم هنا هو عن المعدل النسبي للزيادة في هذين المتغيرين.

نفكر أولاً في الفرضية القائلة بأن الرفاهية تنمو بشكل أبطأ من التكلفة. هذا من شأنه أن يوحي بأن أكبر قدر من الرفاهية الكاملة يمكن الحصول عليه من خلال تشييد أعداد هائلة من العقول الصغيرة. إذا كان هذا صحيحًا ، قد يتجاوز عدد العقول التي هي في حجم الحشرة بدرجة كبيرة عدد السكان من البشر في القدرة الإجمالية على الرفاهية ؛ وفي المستقبل ، سيكون لأعداد

هائلة من العقول الرقمية المؤهلة بالحد الأدنى الأسبقية على كل من العقول التي في حجم الحشرة والكائنات ذات الحجم البشري أو فوق حجم البشر.

إذا فكرنا ، بدلاً من ذلك ، في الفرضية القائلة بأن الرفاهية تنمو بشكل أسرع من التكلفة. قد يشير هذا إلى الاستنتاج المعاكس: أن أكبر قدر من الرفاهية الكاملة يمكن الحصول عليه من خلال تركيز الموارد في عدد قليل من العقول العملاقة.

يبدو أن الحالة التي تكون فيها العقول بمقياس عقول البشر تمثل ، في وضعها الأمثل ، حالة خاصة جدًا ، حيث توجد بعض العقبات الحرجة بخصوص مستوانا أو حيث تتضمن علاقة القياس تعقيدا بخصوص نقطة المقياس البشري. قد تبدو مثل هذه الحالة غير محتملة إلى حد ما من وجهة نظر محايدة ، على الرغم من أنها قد تظهر بشكل طبيعي أكثر في التحليلات التي ترسخ مفهوم الرفاهية في التجربة الإنسانية.

يمكننا أن نسأل بشكل أكثر تحديدًا فيما يتعلق بسمات معينة ، ما إذا كان التعقيد أو العقبة على المستوى البشري أمرًا معقولاً. على سبيل المثال ، يمكننا طرح هذا السؤال حول مقدار الوعي الذي ينشئُه الدماغ. فليس من الواضح على الأقل لماذا يجب أن يكون الأمر متجسدا في الطريقة الأكثر فعالية لتحويل الموارد إلى وعي والتي ينبغي أن تكون من خلال إنشاء عقول بحجم عقل الإنسان ، على الرغم من أنه يجب على المرء أن يفحص نظريات محددة للوعي لإجراء مزيد من البحث في هذه القضية. (13) وبالمثل ، قد يتساءل المرء فيما يتعلق بالوضع الأخلاقي كيف تختلف هذه العقول باختلاف حجم العقل البشري. (14) مرة أخرى ، قد يبدو الادعاء بأن العقول ذات الحجم البشري هي الأمثل في هذا الصدد مريبًا بعض الشيء ، ويفتقر إلى مزيد من التبرير.

حتى لو كان حجم الدماغ البشري هو الأمثل لتوليد الوعي أو الوضع الأخلاقي ، فإن العقول الرقمية ما تزال غير خاضعة للقول بأن بنية الدماغ البشري هي كذلك. إن أجزاء كبيرة من أدمغتنا تبدو غير ذات صلة أو ذات صلة ضعيفة فقط بمقدار الوعي أو درجة الحالة الأخلاقية التي نمتلكها. على سبيل المثال ، يتم تخصيص الكثير من الأنسجة القشرية لمعالجة المعلومات البصرية عالية الدقة ؛ ومع ذلك ، يبدو أن الأشخاص الذين يعانون من ضبابية

الرؤية وحتى الأشخاص المكفوفين تمامًا قادرون على أن يكونوا مدركين تمامًا ويتمتعون بمكانة أخلاقية عالية مثل أولئك الذين يتمتعون بحدة إبصار مثل أعين النسر.

لذلك يبدو من المعقول تمامًا أن وحوش المنفعة ممكنة من خلال تغيير مقياس العقول ، سواء على أساس أن علاقة القياس بين الموارد والقيمة من غير المرجح أن تكون ذروتها هي حجم العقل البشري ، وأيضًا لأن مساحات كبيرة من العقل البشري تبدو ذات صلة منخفضة بدرجة الوعي ، و الحالة الأخلاقية ، أو السمات الأخرى الأكثر ارتباطًا بشكل مباشر بمقدار الرفاهية أو مقدار الرفاهية المساوي للحالة الأخلاقية الناتجة عنها.

3- الآثار الأخلاقية والسياسية لوحوش المنفعة الرقمية المحتملة

Moral and political implications of possible digital utility monsters

دعونا نلخص أبعاد وحوش المنفعة المحتملة التي حددناها:

بعض أبعاد وحوش المنفعة
● القدرة الإنجابية ● تكلفة المعيشة ● السرعة الذاتية ● منحنى المتعة ● نطاق المتعة ● تفضيلات غير مكلفة ● قوة التفضيل ● سلع القائمة الموضوعية وازدهارها ● مقياس العقل

بعض هذه الأبعاد من حالة وحوش المنفعة المحتملة ذات صلة فقط بتحليلات رفاهية معينة. إن إمكانية قوة التفضيل القصوى ، على سبيل المثال ، ذات صلة مباشرة بالتحليلات القائمة على تفضيل الرفاهية ، ولكن ليست مرتبطة بالتحليلات التي تعتمد على المتعة. البعض الآخر ، مثل تكلفة المعيشة ، أكثر صلة بشكل عام بظروف الندرة ، ويبدو أنها تنطبق تقريبا

على أي وجهة نظر تمنح العقول الرقمية الوضع الأخلاقي وتأخذ في الاعتبار التكاليف عند اتخاذ القرارات في هذه الظروف. تختلف الأبعاد أيضًا إلى حد ما في درجات وحوش المنفعة التي يمكن تحقيقها من خلالها ، وفي مدى سهولة وتكلفة الوصول إلى هذه القيم المتطرفة. ومع ذلك ، إذا نظرنا إليها بشكل جماعي ، فإنها تقدم حالة قوية إلى حد ما مفادها أن وحوش المنفعة سوف تصبح مجدية بالفعل عند النضج التكنولوجي. بعبارة أخرى ، سوف يكون الأمر ، وفقًا لمجموعة واسعة من نظريات الرفاهية الشائعة ، هو أن رفاهية أكبر بكثير لكل وحدة من الموارد يمكن تحقيقها من خلال استثمار هذه الموارد في العقول الرقمية بدلاً من البشر البيولوجيين.

لذلك ينشأ سؤالان على درجة كبيرة من الأهمية (يمكننا طرحهما بشكل منفصل عن النظريات الأخلاقية المختلفة):

- كيف ينبغي أن ننظر إلى احتمالية القدرة على إنشاء وحوش المنفعة في المستقبل؟
- كيف ينبغي أن تكون استجابتنا إذا عُرض علينا الأمر واقعًا ، حيث ظهر عدد كبير من وحوش المنفعة؟

3-1- إبداع وحوش المنفعة **Creating utility monsters**

العديد من الآراء ، التي ترى أن إنشاء حياة جيدة جديدة باعتبارها قيمة مهمة ، سوف تنظر إلى احتمال ملء المستقبل بوحوش المنفعة باعتباره أمرًا جذابًا للغاية ، وترى أن الفشل في الاستفادة من هذه الفرصة ، كشيء من شأنه أن يختصر قيمة المستقبل بشكل كبير ، يعد كارثة وجودية. (Bostrom, 2013)

من ناحية أخرى ، يمكن للمرء أيضًا أن يقول بأن لدينا سببًا لعدم إنشاء وحوش المنفعة على وجه التحديد على أساس أنه بمجرد وجود مثل هذه الكائنات ، سوف يكون لها مطالبة سائدة بالموارد النادرة ، وبالتالي سنكون مضطرين لنقل (ربما جميع) الموارد بعيدًا عن البشر إلى وحوش المنفعة هذه ، على حساب الإنسانية. نيكولاس أجار (Nicholas Agar, 2010) ، على سبيل المثال ، يقدم حجة في هذا الاتجاه تمنحنا (على الأقل نسبيًا للإنسان) سببًا أخلاقيًا لمعارضة خلق "ما بعد البشر posthumans" مع مزيج من الوضع الأخلاقي والقوة والرفاهية المحتملة بدرجة أكبر.

ولتبرير هذا الإنكار للرغبة الأخلاقية في إنشاء وحوش المنفعة ، يمكن للمرء أن يستدعي مبدأ "التأثير على الفرد" بما يتماشى مع شعار نارفيسون (Narveson, 1973) : "الأخلاق تتعلق بإسعاد البشر ، وليس بصنع بشر سعداء". (15) إذا كانت واجباتنا هي نحو البشر الموجودين فقط ، وليس لدينا سبب أخلاقي لصنع أشخاص جدد إضافيين ، إذن ، وعلى وجه الخصوص ، لن يكون لدينا أي واجب يدعونا لإبداع وحوش المنفعة ؛ وإذا كان صنع مثل هذه الوحوش يضر بالبشر الموجودين ، فمن واجبنا ألا نصنعها. من المفترض أنه لن يتحتم علينا تجنب إبداعها إذا كان البشر الذين سوف يتضررون من خلقها ينتمون إلى جيل ما في المستقبل بحيث تتشوش هويتهم بسبب اختياراتنا ؛ مع ذلك فإننا على الأقل لن يكون علينا أي واجب إيجابي لإنشائها.

المبدأ الصارم الخاص بالتأثير على الفرد له بعض النتائج غير البديهية. قد يعني ، على سبيل المثال ، أنه ليس لدينا سببا أخلاقيا لاتخاذ أي إجراءات الآن من أجل التخفيف من تأثير تغير المناخ على الأجيال القادمة ؛ وأنه إذا فرضت الإجراءات تكلفة على الجيل الحالي ، فقد يكون لدينا سبب أخلاقي لعدم القيام بذلك. ولأن لها هذه الآثار ، فإن الغالبية سوف يرفضون الأخلاق الصارمة التي تخص التأثير على الفرد. ربما يكون للرؤى الأضعف أو المصنفة جاذبية أوسع. قد يمنَح المرء ، على سبيل المثال ، بعض الوزن الإضافي ولكن ليست الهيمنة الصارمة لصالح الأشخاص الحاليين. وبالمثل ، فإن عدم اليقين الأخلاقي بشأن أخلاقيات البشر يمكن أن يخلِّف حياة جيدة للغاية ذات "قيمة أخلاقية متوقعة" كبيرة أو "جديرة بالاختيار" ، حتى لو كان المرء غير متأكد من قيمة خلق حياة جيدة جديدة.(Greaves and Ord, 2017)

علاوة على ذلك ، تسمح الآراء غير المتكافئة الخاصة بالتأثير على الفرد بالاهتمام الأخلاقي بخصوص التسبب في وجود حياة سيئة – حياة لا تستحق العيش. (Frick, 2014). تؤكد هذه الآراء أن لدينا أسبابًا قوية لتجنب تصنيع عقول رقمية ذات رفاهية سلبية هائلة (وحوش المنفعة السلبية negative utility monsters) وأنه يجب علينا أن نكون على استعداد لقبول تحمل تكاليف كبيرة على السكان الحاليين إذا لزم الأمر لتجنب مثل هذه النتائج. الرؤى الأخرى من وجهات النظر غير المتكافئة ، بينما تنكر أن لدينا أسبابا أخلاقية لملء المستقبل بكائنات جديدة لتجربة أكبر قدر ممكن من الفائدة الإيجابية ، إلا أنها تؤكد أننا ، مع ذلك ، لدينا التزام

أخلاقي بضمان أن تكون المنفعة الخالصة للمستقبل أعلى من خط الصفر. وبالتالي ربما تعلق هذه الآراء أهمية كبيرة على إنشاء ما يكفي من وحوش المنفعة الإيجابية من أجل "تعويض" ما يمكن أن يظهر من وحوش المنفعة السلبية. (Thomas, 2019)

3-2- مشاركة العالم مع وحوش المنفعة Sharing the world with utility monsters

إذا أخذنا في الاعتبار الحالة التي دخلت فيها وحوش المنفعة إلى الوجود بالفعل ، فإن التعقيدات الناشئة عن مبادئ التأثير على الفرد تختفي. من منظور عواقبي consequentialist بسيط ، تكون النتيجة واضحة ومباشرة: يجب علينا نقل جميع الموارد إلى وحوش المنفعة ، وأن نترك البشرية تهلك إذ لم نعد مفيدين من الناحية الأداتية instrumentally.

هناك ، بالطبع ، العديد من الرؤى الأخلاقية التي تنكر أننا ملزمون بنقل جميع مواردنا (ناهيك عن موارد الآخرين) إلى أي كائن سوف يحصل على أكبر قدر من الرفاهية. نظريات علم الأخلاق ، على سبيل المثال ، غالبًا ما تعتبر مثل هذه الأفعال بمثابة تجاوزات كبيرة في حالة التخلي عن ممتلكاتنا ، وغير مسموح بها في مسألة إعادة توزيع ممتلكات الغير.

ومع ذلك ، فإن المبادئ المقبولة على نطاق واسع مثل مدفوعات التحويل الشمولية والمساواة السياسية قد تكون كافية بالفعل لتقديم مقايضات جادة. ضع في اعتبارك الاقتراح الشائع بخصوص الدخل الأساسي العالمي ، الممول من الضرائب ، لتعويض البطالة البشرية التي يسببها الذكاء الاصطناعي المتقدم. إذا كانت مجموعات العقول الرقمية التي تتكاثر بسرعة لديها مطالبة قوية بالدخل الأساسي على الأقل كما يفعل البشر البيولوجيون ، فيمكن عندئذٍ استنفاد القدرة المالية بسرعة. يجب أن ينخفض الراتب المتساوي إلى ما دون مستوى الكفاف البشري (باتجاه مستوى الكفاف للعقل الرقمي) ، في حين أن الراتب غير المتكافئ ، حيث يتم تقنين الدخل على أساس المنافع المتساوية ، سوف يوجه المدفوعات نحو العقول الرقمية ذات التكاليف المعيشية المنخفضة – وبالتالي يمنح عامًا من الحياة لعقل رقمي بدلاً من يوم واحد للإنسان.

قد يبدو أن تجنب هذه النتيجة يتطلب مزيجًا من المعاملة غير الشمولية ، حيث يتم تفضيل البشر المتميزين على العقول الرقمية التي لديها على الأقل مكانة أخلاقية متساوية

ومتطلبات أكبر ، وقيود على فرص التكاثر للعقول الرقمية – وهي قيود ، إذا تم تطبيقها على البشر ، فإن من شأنها أن تنتهك حقوق الإنسان.

وبالمثل ، على المستوى السياسي ، يبدو أن المبادئ الديمقراطية تتطلب أن يكون للعقول الرقمية غزيرة الإنتاج التي تعطي أغلبية هائلة من أعدادها الحق في السيطرة السياسية ، بما في ذلك التحكم في مدفوعات التحويل ونظام حقوق الملكية. (16)

يمكن للمرء هنا أن يسلك الطريق لمحاولة الدفاع عن امتياز خاص للبشر. ربما تقترح بعض النظريات التعاقدية ، على سبيل المثال ، أنه إذا كان البشر في موقع قوة عظمى بالنسبة للعقول الرقمية ، فإن هذا سوف يؤهلنا للحصول على حصة كبيرة من الموارد. بدلاً من ذلك ، يمكن للمرء أن يتبنى تحليلا يستند إلى الاعتبارات المتعلقة بالفرد والتي بموجبها يحق للمجتمعات أو الأنواع أن تميز أعضاءها على الغرباء الذين يتمتعون بأهلية ووضع أخلاقي موضوعي متساوي وبدرجة كبيرة. (17) يبدو أن مثل هذه النسبية تعكس النهج الواقعي الذي تتبعه الدول اليوم ، والتي تعتبر عمومًا أكثر كرمًا في توفير الرعاية لمواطنيها منها تجاه الأجانب ، حتى عندما يكون هناك أجانب أكثر فقرا ، يمكنهم أن يستفيدوا بدرجة أكبر ، ومن حيث الخصائص المتأصلة يستحقون المساعدة مثل مواطني البلد نفسه على الأقل.

قبل السير في هذا الطريق ، يجب على المرء أن يفكر مليًا ونقديًا في السجل التاريخي للمواقف المماثلة التي تم تبنيها على نطاق واسع ولكنها فقدت مصداقيتها منذ ذلك الحين ، والتي تم استخدامها لتبرير اضطهاد العديد من الجماعات البشرية وإساءة معاملة الحيوانات غير الآدمية. سوف نحتاج لأن نسأل ، على سبيل المثال ، ما إذا كانت الدعوة إلى التمييز بين العقول الرقمية والبشر ستكون أقرب إلى اعتناق مذهب من مذاهب التفوق العِرقي؟

نقطة واحدة يجب وضعها في الاعتبار ، هنا ، هي أن العقول الرقمية تكون متنوعة بدرجة كبيرة. بحيث أن بعضها يكون مختلفًا عن الأنواع الأخرى أكثر من اختلاف عقل الإنسان بالنسبة لعقل القطة. إذا تم تشكيل العقل الرقمي بشكل مختلف تمامًا عن العقول البشرية ، فلن يكون من المدهش أن تختلف واجباتنا الأخلاقية تجاهه عن واجباتنا تجاه غيرنا من البشر ؛ وبالتالي فإن معاملته بشكل مختلف لا يجب أن تنتهك أيًا من مبادئ عدم التمييز. بالطبع ، لا

تنطبق هذه النقطة على العقول الرقمية التي تشبه إلى حد بعيد العقول البشرية البيولوجية (مثل محاكاة الدماغ بالكامل) . كما أنها لا تبرر التمييز السلبي ضد العقول الرقمية التي تختلف عن عقول البشر بطرق تمنحها مكانة أخلاقية أكبر أو تجعل احتياجاتها أكثر أهمية أخلاقياً من احتياجات البشر (مثل وحوش المنفعة). وأيضا ، في هذا الصدد ، لن تبرر معاملة العقول الرقمية ذات القدرات أو الإحساس المتشابهة مع المخلوقات غير البشرية وفقًا لنموذج تفاعلاتنا الحالية مع الحيوانات ، حيث تتميز الأخيرة بانتهاكات مروعة وواسعة الانتشار.

تتمثل إحدى طرق محاولة تبرير المعاملة المتميزة للبشر ، دون افتراض تحيز غفل شبيه بالعنصرية لصالح نوعنا كبشر ، في التذرع بمبدأ معين يحق لنا بموجبه (أو نلتزم) بإيلاء اهتمام أكبر للكائنات التي هي أكثر اندماجًا في مجتمعاتنا وحياتنا الاجتماعية من الكائنات الغريبة البعيدة عن نوعنا. يُفترض أن مثل هذا المبدأ مطلوب إذا كان المرء يرغب في إضفاء الشرعية على الطريقة (غير العالمية) التي يُقصِر معظم الناس ومعظم الدول بناء عليها حاليًا معظم المساعدات على أولئك الذين يرون أنهم أعضاء في مجموعاتهم الخاصة. (18) ومع ذلك ، فإن مثل هذه الخطوة لن تستبعد العقول الرقمية التي أصبحت جزءً من نسيجنا الاجتماعي ، على سبيل المثال ، من خلال شغل أدوار مثل ، الإداريين ، أو المستشارين ، أو عمال المصانع ، أو المساعدين الشخصيين. ربما نكون أكثر ارتباطًا اجتماعيًا بمنتجات الذكاء الاصطناعي هذه أكثر من ارتباطنا بالبشر على الجانب الآخر من العالم الذين لم نتواصل معهم أبدًا.

4- مناقشة Discussion

لقد رأينا أن هناك الكثير من الأبعاد المحتملة لوحوش المنفعة في العقول الرقمية. هذا يجعل الاستنتاج بأن وحوش المنفعة ممكنة بالفعل يبدو قويا إلى حد ما. إنه تأثير ضمني لمعظم وجهات النظر الشائعة حاليًا عن الرفاهية.

ما يعنيه هذا هو أنه على المدى الطويل ، سوف تكون الرفاهية الكاملة أكبر بكثير إلى درجة أن تسكن العالم وحوش المنفعة الرقمية بدلاً من الحياة كما نعرفها. وبقدر ما تظهر هذه الكائنات إلى الوجود ، بقدر ما يحتمل أن يسود الصراع بين اهتماماتها من الناحية الأخلاقية واهتمامات الإنسان والحيوان ، على سبيل المثال ، في فترة ندرة الموارد الطبيعية.

ومع ذلك ، في حين أن التركيز المتطرف إما على رفاهية البشرية الحالية أو ، بدلاً من ذلك ، على رفاهية العقول الرقمية الجديدة ، يمكن أن تكون له عواقب وخيمة على الآخر ، سوف يصبح من الممكن لسياسات التسوية أن تعمل بشكل جيد للغاية وفقًا لكلا المعيارين. ضع في اعتبارك ثلاث سياسات ممكنة:

(أ) 100٪ من الموارد للبشر.

(ب) 100٪ من الموارد لوحوش المنفعة.

(ج) 99.99٪ من الموارد لوحوش المنفعة ؛ 0.01٪ للبشر.

من منظور نفعي إجمالي ، تعد (ج) التي تساوي 99.99٪ تقريبًا خيارا جيدا بنفس درجة الخيار المفضل (ب). ومن منظور بشري عادي ، قد تكون (ج) أيضًا مرغوبة أكثر من 90٪ بنفس درجة الخيار المفضل (أ) ، نظرًا للثروة الفلكية التي توفرها العقول الرقمية ، فإن الكثير من الأنظمة يصل حجمها إلى أن تكون أكبر من الإجماليات الحالية. (Bostrom, 2003 ; Hanson, 2001) . وبالتالي ، مسبقًا ، يبدو أمرا جذابا أن نقلل احتمالية كل من (أ) و (ب) في مقابل احتمال أكبر لـ (ج) – سواء كان للتحوط ضد الخطأ الأخلاقي ، أن يصير انعكاسا للتعددية الأخلاقية بشكل مناسب ، أو مراعاة الاعتبارات النظرية للعملية ، أو ببساطة باعتبارها مسألة سياسة واقعية. وكذلك ، حيث يمكن للبشرية أن تزدهر دون إنتاج وحوش المنفعة السلبية، وحيث أن تجنب مثل هذا البؤس هو أمر بالغ الأهمية ليس فقط من وجهات النظر النفعية الشمولية ولكن أيضًا بالنسبة للكثير من وجهات النظر التقييمية الأخرى ، كما أن التدابير التي تقلل من احتمالية وجود وحوش المنفعة السلبية (حتى على حساب البشر) ستكون جزءًا مرحبًا به من حل وسط يمكن أن يتطلب دعما من القاعدة العريضة.

لا يكمن الجزء الصعب من التحدي في وصف مستقبل محتمل تؤدي فيه الإنسانية ومجموع العقول الرقمية أداءً جيدًا للغاية ، ولكن تحقيق نظام يتجنب بصفة دائمة موقفا واحدا من سحق الآخَر، كما نوقش في القسم 3–2.

هذا الجزء من التحدي ينطوي على جانب عملي وأخلاقي. من الناحية العملية ، تكمن المشكلة في ابتكار وسائل مؤسسية أو غيرها من الوسائل التي يمكن بموجبها الحفاظ على

سياسة تحمي أرواح واستحقاقات البشر والحيوانات إلى أجل غير مسمى ، حتى عندما يتفوق على عدد المستفيدين منها وتتفوق عليهم مجموعة كبيرة ومتنوعة من الآلات الذكية عالية القدرة. قد يكون أحد الأساليب لحل هذه المشكلة هو خلق أغلبية ساحقة من العقول الرقمية ذات الرفاهية العالية التي تحفزها على الحفاظ على هذه النتيجة والتي تدعم المعايير والمؤسسات ذات الصلة (متضمَّنة في تصميم الأجيال المتعاقبة من العقول الرقمية).

ومن الناحية الأخلاقية ، فإن السؤال هو ما إذا كانت المقاييس التي أوصت بها تسوية عادلة مسبقة مسموح بها في التنفيذ اللاحق. فقد يكون أحد الاختبارات المفيدة هنا هو ما إذا كان بإمكاننا تأييد تطبيقها على العقول غير الرقمية في ظروف مماثلة. على سبيل المثال ، قد نطلب أن يتوافق أي ترتيب مقترح مع أحد مبادئ عدم التمييز ، كالمثال التالي (Bostrom and Yudkowsky, 2014) :

مبدأ عدم التمييز في الأساس *Principle of Substrate Non-Discrimination*

إذا كان لكائنين نفس الوظيفة ونفس التجربة الواعية ، ولا يختلفان إلا في أساس تنفيذهما ، فعندئذ يكون لهما نفس الوضع الأخلاقي.

و

مبدأ عدم التمييز في تطور الجنين *Principle of Ontogeny Non-Discrimination*

إذا كان لكائنين نفس الوظيفة ونفس التجربة الوعية ، ويختلفان فقط في كيفية ظهورهما في الوجود ، فإنهما يتمتعان بنفس الحالة الأخلاقية.

عند تطبيق هذين المبدأين ، من المهم أن نتذكر النقطة السابقة التي مفادها أن عقول الآلة يمكن أن تكون مختلفة تمامًا عن عقول البشر ، بما في ذلك الطرق التي تخص الكيفية التي يجب أن تُعامل بها. حتى لو قبلنا مبادئ عدم التمييز مثل تلك المذكورة ، يجب أن نتوخى الحذر عند تطبيقها على العقول الرقمية التي ليست نسخًا طبق الأصل من عقل بشري ما.

إذا تأملنا التكاثر ، على سبيل المثال ؛ إذا كان البشر قادرين ، عن طريق صب حطام الحدائق في مفاعل كيميائي حيوي ، على إنجاب طفل كل بضع دقائق ، عندئذٍ يبدو من المرجح

أن المجتمعات البشرية سوف تغير الممارسات القانونية الحالية وتفرض قيودًا على المعدل الذي يُسمح للناس به بالتكاثر. إن عدم القيام بذلك على المدى القصير يؤدي إلى إفلاس أي نظام رعاية اجتماعية ، مع افتراض أن هناك على الأقل بعض الأشخاص الذين قد يرغبون في خلق أعداد هائلة من الأطفال بهذه الطريقة ، على الرغم من افتقارهم إلى وسائل دعمهم. هذا التنظيم يمكن أن يتخذ أشكالًا مختلفة – فقد يُطلب من الآباء المحتملين تقديم سند ملائمة لتلبية احتياجات الأبناء قبل تخليقهم ، أو قد يتم تخصيص تصاريح الإنجاب على أساس هذا الاستحقاق. وبالمثل ، إذا كان لدى البشر القدرة على إنتاج أعداد عشوائية من النسخ الدقيقة لأنفسهم ، فقد نتوقع أن تكون هناك تعديلات دستورية لمنع حسم المنافسات السياسية على أساس من هو الراغب والقادر على إنشاء أكبر عدد من التصويتات. – النسخ. مرة أخرى ، التعديلات يمكن أن تتخذ أشكالًا مختلفة – على سبيل المثال ، قد يضطر من يقوم بتخليق مثل هذه النسخ المكررة إلى مشاركة قوة التصويت الخاصة به مع النسخ التي ينشئُها.

وبالتالي ، بقدر ما تكون هذه التعديلات القانونية أو الدستورية مقبولة للبشر إذا كانت لدينا هذه الأنواع من القدرات الإنجابية ، فقد يكون من المقبول أيضًا إجراء تعديلات مماثلة لاستيعاب العقول الرقمية التي لديها مثل هذه القدرات بالفعل.

السؤال الرئيسي – بالتأكيد من منظور الحياة القائمة – هو ما إذا كان من المسموح أخلاقيا هندسة عقول جديدة لتكون داعمة بشكل موثوق لدعم حقوق وامتيازات معينة لشاغلي الوظائف من البشر. لقد اقترحنا في وقت سابق أن مثل هذا الترتيب للحفاظ على حقوق ملكية الإنسان والامتيازات الاجتماعية يمكن الدفاع عنه ، على الأقل باعتباره مسارًا يحترم عدم اليقين ويخفف من حدة الصراع في التسوية العملية الحكيمة ، سواء كان ذلك هو الأمثل على مستوى النظرية الأخلاقية الأساسية أم لا. قد نشير ، عن طريق القياس ، إلى وجهة النظر الشائعة القائلة بأنه من المقبول أخلاقيا الحفاظ على الأقليات وحمايتها بتكاليف واحتياجات دعم باهظة ، مثل كبار السن ، والمعوَّقين ، ووحيد القرن الأبيض ، والعائلة المالكة البريطانية. يبدو أن هذا الاستنتاج يتعزز إذا افترضنا أن العقول الرقمية التي تم إنشاؤها سوف تؤيد هي ذاتها هذا النظام وتفضل استمراره.

حتى لو كانت النتيجة ذاتها مسموح بها أخلاقياً ، إلا أننا نواجه سؤالاً أخلاقياً آخر ، وهو ما إذا كان هناك شيء غير مرغوب فيه من الناحية الإجرائية من جانب الدقة الهندسية لتفضيلات العقول الرقمية الجديدة التي ننشؤها لضمان موافقتها. يمكننا أن ننظر إلى هذا السؤال من خلال عدسة مبادئ عدم التمييز والنظر في كيفية عرض اقتراح لتطبيق نهج مماثل لتخليق كائنات بشرية جديدة.

وبينما تحاول الثقافات البشرية بشكل منتظم من خلال التعليم ، والحوار ، والتوجيه أن تمرر الأعراف والقيم إلى الأطفال – بما في ذلك طاعة الوالدين واحترام الأعراف والمؤسسات القائمة – يبدو من المحتمل أن اقتراح غرس سلوكيات محددة عن طريق الهندسة الوراثية أمرًا مثيرًا للجدل. حتى إذا وضعنا جانباً الاهتمامات العملية بشأن السلامة ، أو عدم المساواة في الوصول إلى المعلومات أو المناصب ، أو سوء المعاملة من قبل الحكومات القمعية ، أو اتخاذ الآباء خيارات ضيقة الأفق أو حمقاء من جهة أخرى ، فقد يظل هناك قلق من ممارسة السيطرة الجلية على ميول النسل، خاصة إذا تم ذلك بواسطة "عقلية هندسية" واستخدام أساليب تتجاوز تمامًا عقل وإرادة الفرد الخاضع للرقابة (بواسطة تلك التي تجري قبل ولادة الفرد) والتي سوف تكون بطبيعتها إشكالية أخلاقية. (19)

وحيث لا يمكننا تقييم هذه المخاوف بشكل كامل هنا ، فمن الجدير بالذكر أن هناك اختلافين مهمين في حالة العقول الرقمية. الأول هو أنه ، على عكس التكاثر البشري ، قد لا يكون هناك "نموذج افتراضي" واضح يمكن للمبدعين الالتزام به. قد يتخذ المبرمجون قرارات حتمية عند إنشاء ذكاء آلي – سواء لإنشائه بطريقة أو بأخرى ، سواء تدريبه على هذا الهدف أو ذاك ، سواء إعطائه مجموعة تفضيلات أو أخرى. بالنظر إلى أنه يتعين عليها اتخاذ مثل هذا الاختيار ، فقد يعتقد المرء أنه من المعقول أن يتخذ خيارًا يؤدي إلى المزيد من النتائج المرغوبة. ثانيًا ، في حالة الإنسان "المهندَس engineered" بحيث تكون لديه مجموعة معينة من الرغبات، ربما نشك في أنه ربما تظل ، على مستوى أعمق ، تصرفات وميول أخرى قد يتعارض معها التفضيل الهندسي. من المحتمل أن نشعر بالقلق ، على سبيل المثال ، من أن النتيجة يمكن أن يشعر الفرد بالذنب وبقوة بسبب خيبة أمل والديه وبالتالي يضحّي باهتمامات أخرى وبشكل مفرط ، أو أن بعض الأجزاء المخفية من مكوناته النفسية سوف تظل مكبوتة ومُحبطة.

على الجانب الآخر ، في حالة العقول الرقمية ، قد يكون تجنب مثل هذه المشكلات ممكنًا ، إذا كان من الممكن تصميمها بحيث تكون أكثر اتساقًا داخليًا ، أو إذا تمت إضافة تفضيل احترام مصالح السكان "الشرعيين" في "لمسة خفيفة" وبطريقة لا تخلق صراعًا داخليًا ولا تعيق قدرة العقل الرقمي على ممارسة أعماله الأخرى.

بشكل عام ، نظرًا للمخاطر الكبيرة والإمكانية النظرية لتحقيق نتائج عالية جدًا من منظور تقييمي غير شخصي ومن منظور تقييمي محوره الإنسان ، يبدو أن هذا الموضوع يستحق بذل جهود بحثية إضافية لاكتشاف المسارات الممكنة عمليًا والمقبولة أخلاقياً والتي من شأنها تمكين إنشاء وحوش المنفعة الرقمية والحفاظ على ازدهار سكاني كبير.

الهوامش Footnotes

(*) (2020) ، [مسودة الإصدار 1.1] ، معهد مستقبل الإنسانية ، جامعة أكسفورد.

(**) بروفيسور نيك بوستروم ، مدير معهد مستقبل الإنسانية ، جامعة أكسفورد. كارل شولمان باحث مشارك في ذات المعهد. (صفحات معهد مستقبل الإنسانية، جامعة أكسفورد).

https://www.fhi.ox.ac.uk

https://www.fhi.ox.ac.uk/team/carl-shulman/

1- For helpful comments, we're grateful to Guy Kahane, Matthew van der Merwe, Hazem Zohny, Max Daniel, Lukas Finnveden, Lukas Gloor, Uli Alskelung Von Hornbol, and Luke Muehlhauser.

2- Bostrom (2008a; 2008b)

3- للاطلاع على مناقشة حالة وعي الذكاء الاصطناعي ، انظر Chalmers (2010) . أيضا Kagan (2019) يناقش قضية الوضع الأخلاقي اللاواعي ولكن السريع في الذكاء الاصطناعي.

4- قد يكون من غير الواضح ، على أي حال ، ما إذا كانت نسخة دقيقة أو شبه دقيقة من عقل موجود سوف تشكل شخصًا مميزًا جديدًا أو بدلاً من ذلك بمثابة إنشاء مثيل إضافي للشخص الذي كان عقله بمثابة القالب.

5- من الملائم أيضًا "نقل" ملف من موقع في ذاكرة إلى آخر ، وهي عملية تتضمن عادةً إنشاء نسخة من الملف في الموقع الجديد ، والذي يتعايش لفترة وجيزة مع الأصل قبل حذف هذا الأخير. إذا كان الملف المنقول هو الملف القابل للتنفيذ قيد التشغيل للعقل ، فإن هذا الإجراء سوف يتوافق تقريبًا مع سيناريو النقل الآني. الفلاسفة منقسمون حول ما إذا كان الفرد ينجو من مثل هذه العملية (Bourget and Chalmers, 2014) . إذا كانت العملية تنطوي على نوع من "الوفاة" ذات الأهمية الأخلاقية ، فإن عمليات نقل الملفات هذه ستشكل طريقة أخرى تتعرض فيها العقول الرقمية لإمكانية إلغاء متوسط العمر المتوقع وبشكل جذري.

6- يقول هانسون (2016) في الصفحات 63 – 65 بأن الزيادات في التكلفة مع السرعة سوف تكون مطردة في البداية ، أي أن السرعة X2 تتطلب ما يقرب من X2 من ميزانية الأجهزة ، حتى تصل إلى سرعات خارقة إلى حد كبير.

7- يقول ديفيد بيرس (David Pearce, 1995) بأن العقول البيولوجية يمكن هندستها لتعمل على "حالات متدرجة من الرفاهية" بدلا من المدى الحالي للسعادة والألم.

8- Cf. (Agar, 2010, pp. 164-189) .

9- قد يعتقد المرء أن حالة المتعة التي تجذب انتباه العقل بالكامل وتتجاوز جميع الاهتمامات الأخرى سوف تشكل حدًا أقصى من حيث المبدأ من شدة المتعة. ومع ذلك ، يبدو من المعقول أن العقل الأكبر "الأكثر وعيًا" يمكن أن يحتوي بالمعنى المناسب لموضوعنا "قدرًا أكبر" من تجربة المتعة القصوى.

10- As does Parfit (1984, p. 498), citing Rawls (1971, p. 432), who drew from Stace (1944) .

11- لقد وضح (Harsanyi, 1953) أن المقدار المرجح/الموزون من وظائف الاستخدام هو الأمثل في ظل افتراضات معينة ؛ لكن النظرية تترك قيم الأوزان غير محددة.

12- See, e.g., (MacAskill, Cotton-Barratt and Ord, 2020).

13- هذه المسألة لها أهمية خاصة لأن الكثير من نظريات الوعي محددة بشكل كافٍ بأن تعتبر أن التطبيقات الحسابية تبدو عرضة للتطبيقات من أدنى درجة.

(Herzog, Esfeld and Gerstner, 2007) .

14- تقول شيلي كاجان (Shelly Kagan, 2019) ، على سبيل المثال ، بأن الوزن الأخلاقي لمسألة معينة – مثل الاهتمام بتجنب قدر معين من المعاناة – ينبغي ترجيحه بحسب درجة الحالة الأخلاقية للموضوع محل الاهتمام ، حيث تعتمد درجة المكانة على سمات نفسية مختلفة.

15- يقدم فريك (Frick, 2014) محاولة حديثة لتفسير يتماشى مع الموقف العام.

16- Cf. (Calo, 2015).

17- E.g. (Williams, 2006).

18- بالطبع ، هذه الممارسات عرضة للنقد على مستوى واسع ، على سبيل المثال ، نقد بواسطة سينجر (Singer, 1981) وبواسطة أبيا (Appiah, 2006) .

19- E.g. (Habermas, 2003; Sandel, 2007) .

References

- Agar, N. (2010) *Humanity's End* . The MIT Press. pp. 164–189.
- Aghion, P., Jones, B. F. and Jones, C. I. (2017) 'Artificial Intelligence and Economic Growth', *National Bureau of Economic Research Working Paper Series* , No. 23928.
- Appiah, A. (2006) *Cosmopolitanism: Ethics in a World of Strangers* . Allen Lane.
- Bostrom, N. (2003) 'Astronomical Waste: The Opportunity Cost of Delayed Technological Development', *Utilitas* , 15(3), pp. 308–314.
- Bostrom, N. (2008a) 'Letter from Utopia', *Studies in Ethics, Law, and Technology* , 2(1).
- Bostrom, N. (2008b) 'Why I Want to be a Posthuman when I Grow Up', in Gordijn, B. and Chadwick, R. (eds) *Medical Enhancement and Posthumanity* . Springer Netherlands, pp. 107–136.
- Bostrom, N. (2013) 'Existential Risk Prevention as Global Priority', *Global Policy* , 4(1), pp. 15–31.
- Bostrom, N. and Yudkowsky, E. (2014) 'The Ethics of Artificial Intelligence', in Frankish, K. and Ramsey, W. M. (eds) *The Cambridge Handbook of Artificial Intelligence* . Cambridge University Press, pp. 316–334.
- Bourget, D. and Chalmers, D. J. (2014) 'What do Philosophers Believe?', *Philosophical Studies* , 170(3), pp. 465–500.
- Calo, R. (2015) 'Robotics and the Lessons of Cyberlaw', *California Law Review* , p. 529.
- Chalmers, D. (2010) 'The Singularity: A Philosophical Analysis', *Journal of*

Consciousness Studies , 17, pp. 7–65.

- Frick, J. D. (2014) *'Making People Happy, Not Making Happy People': A Defense of the Asymmetry Intuition in Population Ethics (Doctoral dissertation).*

- Greaves, H. and Ord, T. (2017) 'Moral Uncertainty About Population Axiology', *Journal of Ethics and Social Philosophy* , 12(2), pp. 135–167.

- Habermas, J. (2003) *The Future of Human Nature* . Polity Press.

- Hanson, R. (2001) *Economic Growth given Machine Intelligence* . Technical Report, University of California, Berkeley.

- Hanson, R. (2016) *The Age of Em: Work, Love, and Life when Robots Rule the Earth* . Oxford University Press. pp. 63–5.

- Harsanyi, J. (1953) 'Cardinal Utility in Welfare Economics and in the Theory of Risk-taking', *Journal of Political Economy* , 61(5), pp. 434–435.

- Herzog, M. H., Esfeld, M. and Gerstner, W. (2007) 'Consciousness & the Small Network Argument', *Neural Networks: The Official Journal of the International Neural Network Society* , 20(9), pp. 1054–1056.

- Hurka, T. and Tasioulas, J. (2006) 'Games and the Good', *Proceedings of the Aristotelian Society, Supplementary Volumes* , 80, p. 224.

- Kagan, S. (2019) *How to Count Animals, more or less* . Oxford University Press.

- MacAskill, W., Cotton-Barratt, O. and Ord, T. (2020) 'Statistical Normalization Methods in Interpersonal and Intertheoretic Comparisons', *Journal of Philosophy* , 117(2), pp. 61–95.

- Narveson, J. (1973) 'Moral Problems of Population', *The Monist* , *57* (1), pp. 62–86.

- Nordhaus, W. D. (2007) 'Two Centuries of Productivity Growth in Computing', *The Journal of Economic History* , 67(1), pp. 128–159.

- Nozick, R. (1974) *Anarchy, State, and Utopia* . Basic Books. p. 41.

- Parfit, D. (1984) *Reasons and Persons* . Oxford University Press, pp. 343, 388–389, 498.

- Pearce, D. (1995) *Hedonistic Imperative* . www.hedweb.com [accessed: 24 Sept 2020].

- Rawls, J. (1971) *A Theory of Justice* . Belknap. pp. 379–380.

- Sandel, J. M. (2007) *The Case Against Perfection: Ethics in the Age of Genetic Engineering* . Harvard University Press.

- Singer, P. (1981) *The Expanding Circle: Ethics and Sociobiology* . Clarendon Press.

- Stace, W. T. (1944) 'Interestingness', *Philosophy* , 19(74), pp. 233–241.
- Thomas, T. (2019) 'Asymmetry, Uncertainty, and the Long term', GPI Working Paper No. 11–2019.
- Williams, B. A. O. (2006) 'The Human Prejudice', in *Philosophy as a humanistic discipline* . Princeton University Press. pp. 135–152.