

人工道德

文·圖／吳誠文

「少爺，您現在不能進這個房間。」
「為什麼？我只是進去找個東西。」
「老爺交代，今天不能讓任何人進這個房間。」
「臭機器人，不要煩我！」
「對不起，少爺，我不能讓您進去。」
「你滾開！」
「老爺的命令必須執行，請馬上離開，不然您可能會受傷。」
「你敢？你忘了機器人的戒律了？」
「老爺的命令必須執行，請馬上離開。」

人類社會秩序的維持，人群的和睦相處，主要靠的不是法律，而多半是道德（這樣講如果冒犯了學法律的您，請原諒我的偏見），因為法律與道德（情理）無法完全契合，也永遠跟不上社會變遷的腳步，而道德是行為能不能被社群接受的不變（或緩慢調整）的基準。不只人的社會有道德基準，許多其他群居動物也可能有他們自己的道德基準（所以人有人倫，狗有狗倫，不能一概而論）。可是你會說，如果有一個村子，裡頭的人完全沒有道德的觀念，私慾凌駕一切，那是不是只有法律可以維持秩序，讓大家和平相處？但是即使是這樣，法律的制定應該是根源於人群的共同意見與需求，亦即要符合道德的觀念，所以這是矛盾而不可能的，因此道德還是最基本的行為規範的基礎。可是有一個問題來了，科技的進步讓我們可以預期不久的未來真的會有一種新的「人」存在於我們的社群裡，這種人叫機器人。機器人具有人工智慧，但是如果它犯錯（相較於人的利益），你要用法律制裁它嗎？還是你要教它道德（就像要狗學習人的道德一樣）？機器人可以理解道德嗎？科幻小說家Isaac Asimov在他1942年的短篇小說《Runaround》裡提出了「機器人三律」（The Three Laws of Robotics）：

1. 機器人不可以傷害人，也要盡量避免人受到傷害。
2. 機器人必須聽從人的命令，除非違反第一條法律。
3. 機器人必須盡量維護自己生命，除非違反第一條或第二條法律。

這三條法律雖然後來廣泛被引用（包括在學術與產業技術研究的領域），不過顯然不夠完備，無法讓機器人跟人和平共處。問題是，法律與道德本就無法完全契合，因此迄今也沒有人可以幫機器人制定完備的法律，保證機器人在人類社會中不會出現與人衝突的狀況。事實上，機器人是人創造的，它的智慧與靈魂是創造者植入它身體裡的軟體。既然

人會刻意傷害其他人或物是千古不變的行為，人類未來可能被機器人消滅的疑慮便一直是存在的（如果機器人比人聰明且強壯，且沒有道德規範的話）。當我們創造人工智慧的同時，我們也能為機器人創造人工道德嗎？人工道德是什麼？是人的道德的一部分還是超出了人的道德範疇？人的道德又是什麼？誰是真正的有德之士（我知道的都是沒有留下瑕疵記錄的古人，而今人只要通過選舉的X光機都保證淘汰出局）？這些問題沒有人能回答得讓所有人滿意，所以人類在擔心步北極熊家破熊亡後塵的同時，也得要擔心鐵奴當道，禍起蕭牆吧。

在棒球場上，有時候有些球員會為了贏球而做一些欺騙對手或故意干擾對手的動作，例如一壘手把球藏起來，跟投手一起演戲，讓對方跑壘者以為球在投手的手套裡，當他不注意而離壘時便被一壘手觸殺出局（我小時候打球便吃過這個虧，當時非常看不起那個騙我的人）。我們看職業棒球比賽時也常常看到從一壘往二壘跑的跑壘者在滑壘時會刻意滑向在二壘準備往一壘傳球的對手（通常是游擊手或二壘手），逼他採取閃躲的動作以延緩他的傳球速度甚至讓他傳球失去準度，希望避免被雙殺。這兩個例子都沒有違反棒球規則（法律），但是在道德上都有模糊空間與爭議性。前者的爭議是欺騙的行為違反運動精神，但是它如何被認定是欺騙？如何在棒球規則裡規範？如何由裁判執法？這是非常困難的。後者雖然不是欺騙的行為，但是危險，如果其中一人不熟悉這種行為，沒有默契的話很容易造成對彼此的傷害（棒球選手穿釘鞋，滑壘時可能傷到對手，對手跳起來也可能踩傷自己）。可是認定是不是故意也不容易，所以在棒球規則裡仍然沒有規範。沒有規範的行為我們就說它是球場道德

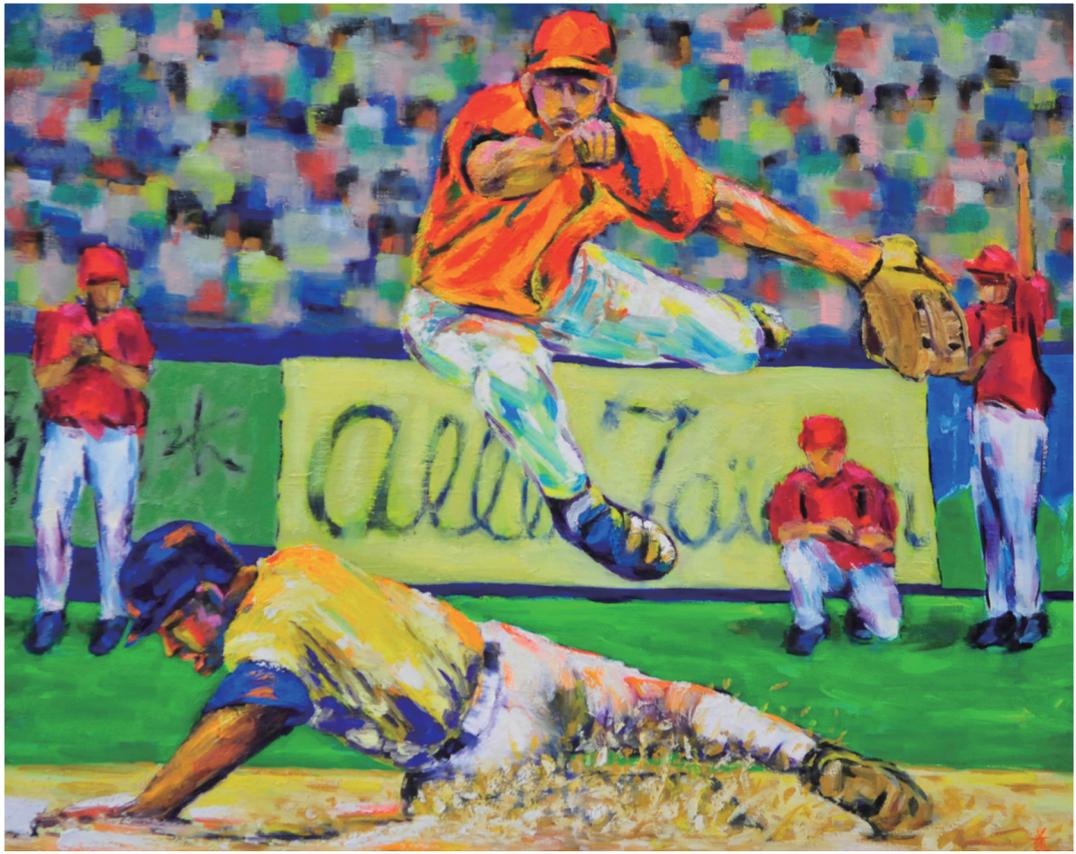


德國慕尼黑啤酒節時，古老的餐廳裡有管樂隊以古老的樂器吹奏著古老的曲調。你能想像未來這個畫面裡的樂手是機器人嗎？在機器人三律的規範下，他們在接到眾多餐廳顧客不同的喜惡要求時會如何的手足無措呢？面對喝醉酒挑釁的客人時他們又要如何自處呢？

與默契的問題吧，那這個尺度怎麼拿捏，可能會因人因時因地而異。一個投手投出打擊者的觸身球是不是故意的，這通常也難以判斷，但是因為它的危險性太大，可能造成打擊者的重大傷害，所以雖然裁判執法時爭議不斷，棒球規則仍不得不規範。相同的法律會因為執法者的道德標準不同而有落差，因此不難理解判決會出現爭議。再者，道德屬於文化的內涵，因此有地域性，並非放諸四海皆準，例如中南美洲的棒球選手跑壘時非常積極而具侵略性，經常會以合法的滑壘動作刻意威脅防守壘包的對方選手的安全（連一壘手都有可能受到滑壘者釘鞋的踩踢），試圖逼對手棄守或失誤，這一點就跟顧慮安全謹守禮儀的亞洲棒球選手有非常明顯的差異，但是他們卻習以為常（從少棒到成棒皆如此）。

如同棒球比賽一樣，學術界爭議性的行為也常常是道德問題而非法律問題，因此學者們逐漸

吳誠文 專欄。



這幅印象畫中，滑壘的跑壘者刻意滑向防守二壘的對方選手，逼他必須跳起閃避，同時在空中把球傳向一壘。由於這種高難度的動作已達藝術境界，受到觀眾喜愛，因此雖然危險，規則並未禁止，有經驗的雙方球員也皆有默契。（油畫提供：馮蕙慈）

了解大家必須一起訂定學術倫理的規範，不能老是歸咎於法律。只是道德的問題往往因人因時因地而異，因此雖有倫理規範，爭議卻從不間斷，違反學術倫理的情形仍層出不窮，社會上其他專業領域的道德問題亦復如是。既然人類的道德問題沒有人能徹底解決（想出面解決的人也可能不是大家認同的有德之士），它所延伸的人工道德與機器人倫理的問題更將是雪上加霜。（2015.12.14）

更正啟事：前期〈是成長是演化是塵土〉第34頁圖說，紅木的英文更正為“Redwood”。



吳誠文小檔案

1971年巨人隊少棒國手，為國家捧回世界少棒冠軍盃。臺南一中畢業後，考進臺大電機系，1981年從臺大電機系畢業，1984年負笈美國深造，1987年取得美國加州大學聖塔芭芭拉校區電機與電腦工程學博士。學成返國任教於清華大學電機系，2000-2003年兼任系主任，2004-2007年擔任電機資訊學院院長。鑽研超大型積體電路設計與測試和半導體記憶體測試，卓然有成，2004年當選IEEE Fellow。2007年借調至工研院主持系統晶片科技中心（STC），2010年將STC整合至資訊與通訊研究所（ICL），並接任所長，2013年獲經濟部國家產業創新獎的最高榮譽，卓越創新研究機構獎。同年獲教育部國家講座主持人榮譽，2014年歸建清華大學擔任副校長。