**หมากล้อม, รถไฟ, รถไร้คนขับ, และปัญญาทางจริยธรรมของมนุษย์ที่ AI คิดแทนไม่ได้**

เหมือนมาด มุกข์ประดิษฐ์

(สาขาปรัชญา มหาวิทยาลัยธรรมศาสตร์)

Abstract

This paper considers the ways in which artifitial intelligence took over many areas previously thought to be essentially human, from strategy planning in the game of Go to composing classical music. In the area of morality, it is thought that humans can program driverless car to make moral decisions. I ask whether we can imagine an AI having practical wisdom as humans do. Upon considering different ways people have formed moral questions, I suggest that we could imagine AI making moral decisions where the questions are formed in terms of a calculation for the best outcome. However, only humans can be said to possess practical wisdom, where practical wisdom means the ability to discern the situations rather than the ability to calculate the best means to an end.

บทคัดย่อ

บทความนี้พิจารณาปัญญาประดิษฐ์ซึ่งเข้ามาหน้าที่คิดแทนมนุษย์ และคิดแทนมนุษย์ได้ในหลายสาขาซึ่งเราไม่เคยคิดว่าปัญญาประดิษฐ์จะเข้ามาแทนที่มนุษย์ได้ เช่นการเล่นหมากล้อมและการทำงานศิลปะ ผู้เขียนพิจารณาคำถามที่ว่าปัญญาประดิษฐ์จะเข้ามาแทนที่การคิดของมนุษย์ในพื้นที่ทางจริยธรรมได้หรือไม่ โดยดูจากกรณีรถไร้คนขับ ผู้เขียนเสนอว่าปัญญาประดิษฐ์จะเข้ามาแทนที่การคิดของมนุษย์ได้ในกรณีที่คำถามทางจริยธรรมถูกทำให้เป็นการคำนวณทางคณิตศาสตร์ ซึ่งคำถามเช่นนั้นไม่ได้สะท้อนความลึกซึ้งและรุ่มรวยของปัญญาทางจริยธรรมของมนุษย์ ผู้เขียนเสนออีกว่าปัญญาประดิษฐ์จะไม่สามารถเข้ามาแทนที่การคิดของมนุษย์ได้ในแง่มุมทางจริยธรรมที่มีความหมาย และปัญญาทางจริยธรรมจึงเป็นสิ่งที่สะท้อนความเป็นมนุษย์ที่ไม่มีสิ่งใดมาแทนที่ได้

1. "The Disruption"

นักวิทยาศาสตร์และนักเขียนเกี่ยวกับประเด็นทางปรัชญา แซม แฮริส (Sam Harris) มีจุดยืนว่าวิชาจริยศาสตร์เป็นวิชาที่ล้าสมัยแล้ว เพราะเขาคิดว่าวิทยาศาสตร์สามารถเข้ามาแทนที่ปรัชญาได้อย่างสมบูรณ์ เนื่องจากคำถามทางจริยศาสตร์เป็นคำถามที่นักวิทยาศาสตร์ตอบได้ดีกว่า ในการดีเบตสำคัญจัดโดยโครงการ The Origins Project ของมหาวิทยาลัยอริโซน่าเสตท ที่ซึ่งแฮริสถกกับนักจริยศาสตร์ที่มีชื่อเสียงอย่าง ปีเตอร์ ซิงเงอร์ (Peter Singer) และ ไซมอน แบลคเบิร์น (Simon Blackburn) เขากล่าวว่าทุกคนรู้อยู่แล้วว่าความทุกข์ทรมานมากที่สุดเป็นเวลานานที่สุดของทุกคนเป็นสิ่งเลวร้าย สิ่งอื่นๆนอกจากนั้นจะเลวร้ายบนสเกลของความห่างจากจุดนั้น คือห่างมากก็ดีมาก ห่างน้อยก็ดีน้อยหรือไม่ดีเลย แฮริสคิดว่านี่เป็นประเด็นที่นักปรัชญาเฝ้าถกเถียงกัน แต่มันเป็นการเถียงที่ไม่ได้ประโยชน์อะไร วิถีของมนุษย์ชาติต่อไปนี้จะเป็นการถามคำถามว่าอะไรจะทำให้ห่างจากจุดเลวมากที่สุดนั้นได้ นี่คือคำถามของวิทยาศาสตร์ แฮริสกล่าวว่า คุณค่าทางจริยธรรมสามารถถูกลดทอนได้เป็นความจริงทางวิทยาศาสตร์ ("values reduce to facts") (Sam Harris, https://youtube.be/9pjOU7WP6KA และโปรดดู Sam Harris. (2011). The Moral Landscape. NY: Free Press.) ในความคิดของแฮริส ปัญญาทางจริยธรรม (practical wisdom หรือ *phronesis* ที่อริสโตเติลเรียก) คือการคำนวณข้อเท็จจริง (facts) เท่านั้น ไม่ใช่เรื่องของการให้คุณค่า จึงเป็นเรื่องที่นักวิทยาศาสตร์ทำแทนนักปรัชญาได้ นักปรัชญาสายประโยชน์นิยม ปีเตอร์ ซิงเงอร์ ซึ่งเป็นนักประโยชน์นิยมในเวทีนั้นแสดงความไม่เห็นด้วยกับแฮริส โดยกล่าวว่า แม้จะเป็นการคำนวณประโยชน์สุข แต่ความเข้าใจชีวิตมนุษย์ที่ทำให้การคำนวณประโยชน์สุขเป็นไปได้ ก็เป็นเรื่องที่จำเป็นต้องใช้ทักษะในทางปรัชญาเพื่อถกเถียงหาข้อสรุป และนี่เป็นคำถามที่สาขาวิชาจริยศาสตร์ประยุกต์ (applied ethics) ถกเถียงเป็นประจำ เช่นว่า ถ้ามีทรัพยากรจำกัด จะขยายอายุขัยคนจำนวนน้อยไปถึงสองร้อยปีแต่มีคุณภาพชีวิตด้อยลง หรือจะไม่ขยายอายุขัยเพื่อรักษาคุณภาพชีวิต และเพื่อกระจายทรัพยากรไปให้คนจำนวนมาก ซิงเงอร์ไม่ได้เถียงแฮริสว่าจริยศาสตร์เป็นเรื่องของการคำนวณหาความสุขมากที่สุด แต่เขาเถียงว่าสิ่งใดจะนับเป็นความสุข ต้องอาศัยทักษะบางอย่างนอกเหนือไปจากทักษะทางเทคนิคของนักวิทยาศาสตร์ วิธีคิดของซิงเงอร์คือ ความรู้ทางวิทยาศาสตร์ไม่อาจบอกได้ถึงหลักการแรกทางจริยศาสตร์ (ความคิดที่พื้นฐานที่สุดทางจริยศาสตร์ เช่นหลักมหสุขของประโยชน์นิยม) และนอกจากนั้นยังไม่อาจบอกได้ว่าอะไรนำไปสู่หลักการแรก (อะไรนับเป็นความสุข เช่นการมีอายุยาว หรือการมีคุณภาพชีวิตดี) ความรู้เช่นนี้มาจากอีกแหล่งหนึ่ง มันเป็นความรู้ทางปรัชญา ไซมอน แบลคเบิร์น นักคิดสายอริสโตเติล นั้นไม่เห็นด้วยกับแฮริสตั้งแต่บอกว่าจริยศาสตร์เป็นเรื่องของการคำนวณแล้ว เขาคิดว่าจริยศาสตร์ไม่ใช่เรื่องของข้อเท็จจริง (facts) แต่เป็นเรื่องของการให้คุณค่า (values) เช่นว่า มนุษย์ทั่วไปไม่คิดว่าความสุขที่ได้มาจากการสมองพิการ มีสมองเหมือนเด็กห้าขวบตลอดชีวิต เป็นสิ่งดี เพราะมนุษย์ให้คุณค่ากับปัญญาและความรู้ จริยศาสตร์ไม่ใช่เป็นการคำนวณ (calculate) แต่เป็นการพิจารณาตัดสิน (deliberate) นั่นคือแบลคเบิร์นบอกว่าวิทยาศาสตร์ช่วยให้ความรู้ทางจริยศาสตร์ไม่ได้ จริยศาสตร์เป็นเรื่องของปรัชญา (และมนุษยศาสตร์ ในฐานะที่เป็นศาสตร์ที่ให้ความเข้าใจเกี่ยวกับความเป็นมนุษย์) ข้อโต้แย้งนี้ให้บริบทแก่เราในการพิจารณาว่าปัญญาของมนุษย์ต่างจากปัญญาประดิษฐ์อย่างไร อีกนัยหนึ่งคือ ปัญญาประดิษฐ์สามารถแทนที่ปัญญาของมนุษย์ได้หรือไม่ในสาระสำคัญ

ปัจจุบันนี้ปัญญาประดิษฐ์สามารถทำอะไรได้หลายอย่าง ที่น่าสนใจคือมันสามารถเรียนรู้เองได้ หมายความว่าเรียนรู้จากความผิดพลาดของตัวเองได้ มีความคิดสร้างสรรค์ได้ ตัวอย่างเช่นมีโปรแกรมการเล่นหมากล้อมจีนชื่อ AlphaGo ซึ่งสามารถเอาชนะแชมป์ที่เป็นมนุษย์ได้ ปรากฏการณ์นี้เป็นที่จับตาของคนที่สนใจเรื่อง AI เป็นพิเศษ เพราะหมากล้อมจีน ถือเป็นเกมที่วัดความฉลาดลึกล้ำของมนุษย์มาแต่โบราณ มันมีความหมายมากกว่าเกม เพราะเป็นการฝึกกลยุทธทางการทหาร เป็นการฝึกปัญญาของบัณฑิตสมัยก่อน ความเป็นไปได้ของการเดินหมากแต่ละตามีมากมายเหลือคณานับ คนที่เป็นแชมป์ในเกมนี้ถือเป็นมนุษย์ที่ฉลาดมากอย่างไม่ต้องสงสัย การที่ AI สามารถคิดได้ดีกว่ามนุษย์ ทำให้เราต้องตั้งข้อสงสัยว่า AI จะแทนที่ปัญญาของมนุษย์ได้อย่างสมบูรณ์หรือไม่ เราอาจคิดว่า AI อาจสามารถเรียนรู้เองได้ แต่ไม่สามารถคิดสร้างสรรค์ได้ (เช่นสร้างงานศิลปะ) และอย่างน้อยปัญญาของมนุษย์ในด้านที่คิดสร้างสรรค์ก็ยังเป็นสิ่งที่ AI แทนที่ไม่ได้ แต่ปรากฏว่ามีโปรแกรมคอมพิวเตอร์ที่แต่งเพลงได้ ชื่อว่า AIVA (Artificial Intelligence Virtual Artist) AIVA สามารถแต่งเพลงคลาสสิค และแต่งได้อย่างไพเราะ คนที่ไม่รู้มาก่อนจะบอกไม่ได้ว่าเพลงนั้นถูกแต่งโดย AI หรือโดยมนุษย์ ปรากฏการณ์เช่นนี้ยิ่งทำให้เราตั้งข้อสงสัยว่า AI สามารถแทนที่ปัญญาของมนุษย์ได้อย่างสมบูรณ์หรือไม่ ข้าพเจ้าต้องการคิดต่อจากความคิดของซิงเงอร์ และโดยเฉพาะอย่างยิ่งต่อจากความคิดของแบลคเบิร์น และโต้แย้งว่า มีอยู่พื้นที่หนึ่งที่ AI ไม่สามารถแทนที่ปัญญาของมนุษย์ได้ คือ AI ไม่สามารถมีปัญญาทางจริยธรรม (practical wisdom) ได้ ข้าพเจ้าต้องการพิจารณาว่าลักษณะของปัญญาทางจริยธรรมคืออะไร เกี่ยวพันกับความเป็นมนุษย์อย่างไร และทำไมสิ่งที่ไม่ใช่มนุษย์ไม่สามารถมีปัญญาชนิดนี้ได้ ข้าพเจ้าเสนอว่าปัญญาของ AI เทียบได้กับความรู้ทางวิทยาศาสตร์ที่แฮริสเสนอว่าจะมาแทนที่ปัญญาทางจริยธรรมได้ เพราะทั้งคู่เป็นการคำนวณข้อเท็จจริง ไม่เกี่ยวกับการให้คุณค่า แบลคเบิร์นโต้แย้งว่าความรู้ที่มาจากการคำนวณข้อเท็จจริงไม่อาจให้ความกระจ่างทางจริยธรรมได้ ข้าพเจ้าจะเสนอว่าปัญญาของ AI ก็ไม่อาจให้ความกระจ่างทางจริยธรรมได้เช่นกัน

2. การ "แก้ปัญหา" (solution) กับ การ "มองปัญหา" (perception)

เมื่อกล่าวถึงความสามารถของ AI ในการตัดสินทางจริยธรรม หลายคนคิดถึงประเด็นเรื่องรถไร้คนขับ (driverless cars) ซึ่งเป็นที่ถกเถียงกันยังหาข้อสรุปไม่ได้ อันที่จริงเทคโนโลยีรถไร้คนขับมีพร้อมแล้ว แต่ประเด็นทางจริยธรรมยังเป็นที่ตกลงกันไม่ได้ อันที่จริงปัญหาทางจริยธรรมของรถไร้คนขับ เป็นปัญหาที่นักจริยศาสตร์รู้จักกันดีในนามการทดลองทางความคิดเรื่อง The Trolley Problem ในการทดลองทางความคิดนี้ ผู้อ่านต้องสร้างสถานการณ์สมมติว่ามีรถไฟกำลังแล่นมาจะชนคนซ่อมรางห้าหกคน ผู้อ่านยืนเห็นเหตุการณ์อยู่ด้านข้าง และสามารถกดสวิทช์เพื่อสลับรางรถไฟได้ ปัญหาคือถ้าสับรางแล้วต้องไปชนคนหนึ่งคนที่ไม่รู้เรื่องรู้ราวอะไรเลย เป็นผู้บริสุทธิ์เหมือนกับช่างซ่อมรางห้าคนนั้น ปัญหาคือ ผู้อ่านจะกดสวิทช์หรือไม่ ปัญหานี้นักจริยศาสตร์อย่างปีเตอร์ ซิงเงอร์ใช้เพื่อจะพิสูจน์ความไม่คงเส้นคงวาของความรู้สึก (sentiments) ทางจริยธรรมของคนทั่วไป เนื่องจากคนทั่วไปส่วนใหญ่จะตัดสินใจกดสวิทช์ แต่ถ้าผู้ถามถามต่อไปว่า ถ้าเปลี่ยนจากการกดสวิทช์ เป็นการต้องผลักคนอ้วนลงมาให้รถไฟชนเพื่อจะหยุดรถไฟและช่วยคนซ่อมรางห้าหกคน ผู้อ่านจะทำไหม (ต้องเป็นคนอ้วนเพื่อป้องกันทางออกอีกทางหนึ่ง คือผู้อ่านจะตัดสินใจกระโดดลงมาหยุดรถไฟเอง) ในกรณีนี้ผู้อ่านหรือผู้ฟังส่วนใหญ่จะบอกว่าไม่ทำ นักประโยชน์นิยมชี้ให้เห็นว่า ความรู้สึก ทางจริยธรรมของคน เป็นสิ่งที่พึ่งพาไม่ได้ในการคิดตัดสินประเด็นทางจริยธรรม เราควรใช้เหตุผลเท่านั้น และเหตุผลในที่นี้คือการคำนวณหาการกระทำที่จะก่อความสุขปริมาณมากที่สุด ปัญหารถไร้คนขับ เกิดขึ้นหลังจากมีปัญหาเรื่อง Trolley Problem แต่เป็นเสมือนการเอา Trolley Problem ซึ่งเป็นการทดลองทางความคิด มาอยู่ในโลกแห่งความเป็นจริง เพราะคนที่เขียนโปรแกรมรถไร้คนขับ ต้องตัดสินใจว่า ในกรณีเกิดอุบัติเหตุ ถ้าต้องเลือกแล้วรถจะเลือกชนใคร จะรักษาชีวิตคนนั่งไว้เป็นสำคัญ หรือเลือกรักษาชีวิตคนจำนวนมากเป็นสำคัญ หรือจะเลือกชนคนที่ทำผิดกฏจราจร (เช่นขับปาดหน้า หรือข้ามถนนไม่ตรงทางม้าลาย) และรักษาชีวิตคนที่ทำถูกกฏจราจร หรือเลือกรักษาชีวิตเด็ก (อาจจะตายง่ายกว่าที่รถจะชนผู้ใหญ่) หรือจะเลือกชนคนเลว (สมมติถ้าบอกได้) ที่เรารู้แน่คือ สถิติบอกว่ารถไร้คนขับจะมีอัตราอุบัติเหตุน้อยกว่ามาก อุบัติเหตุส่วนใหญ่เกิดจากความผิดพลาดของมนุษย์ ทีนี้ถ้าคิดแบบประโยชน์นิยม เราอาจคิดว่าถ้าโดยรวมแล้วสามารถลดอุบัติเหตุได้ ความยุติธรรมในอุบัติเหตุกรณีใดกรณีหนึ่ง ก็ไม่สำคัญเท่ากับความจริงที่ว่าโดยรวมแล้วจะมีคนเจ็บตายน้อยลง แต่คนจำนวนมากคิดว่า ชีวิตคนแต่ละคนมีความสำคัญมากพอที่จะชะลอเทคโนโลยีนี้ไว้จนกว่าเราจะแก้ปัญหานี้ได้ สิ่งที่น่าสนใจประเด็นหนึ่งก็คือ ถ้ารถที่ประสบอุบัติเหตุไม่ใช่เป็นรถไร้คนขับ แต่เป็นรถธรรมดาที่มีคนขับ การตัดสินใจของคนขับก็อาจจะไม่ได้ตัดสินใจได้ดีไปกว่า AI เลย คือมนุษย์อาจจะตัดสินไปตามความตกใจมากกว่าจะเป็นการพิจารณาตัดสินทางจริยธรรมอย่างมีสติและรอบคอบ แต่ในกรณีเช่นนั้นสังคมจะให้อภัยกับคนที่อยู่ในภาวะตื่นตกใจ แต่สังคมจะไม่ให้อภัยถ้ามนุษย์สามารถคิดไว้ล่วงหน้าได้ดังในกรณีของการวางโปรแกรมคอมพิวเตอร์ให้รถไร้คนขับตัดสินใจชนใครคนใดคนหนึ่ง ปัญหาคือในสถานการณ์เช่นที่ว่านี้ไม่มีการตัดสินใจที่ "ถูกต้อง" หรืออย่างน้อยก็ไม่มีการตัดสินใจที่ถูกต้องที่คนเห็นพ้องต้องกันได้ (ดังนั้น Trolley Problem จึงเป็นการทดลองทางความคิดที่ยังเป็นที่กล่าวถึงกันทุกวันนี้)

ประเด็นที่น่าสนใจเกี่ยวกับ Trolley Problem ซึ่งเกี่ยวกับประเด็นที่บทความนี้กำลังถกเถียงอยู่ก็คือ การทดลองทางความคิดเช่นนี้สะท้อนความคิดทางจริยศาสตร์แบบยุคภูมิธรรม (enlightenment) ซึ่งมีฐานคติอยู่ว่าจริยศาสตร์เป็นเรื่องของการใช้เหตุผลแบบไม่เข้าใครออกใคร มีทางออกที่ถูกต้อง คำถามทางจริยธรรมคือคำถามเกี่ยวกับการกระทำที่ถูกต้อง ซึ่งมีอยู่แบบเดียวในแต่ละสถานการณ์ และในแต่ละสถานการณ์มีกฏหรือบรรทัดฐานทางจริยธรรมแบบเดียวที่ใช้ได้ทุกที่ทุกเวลา (เปรียบเทียบกับความคิดแบบอริสโตเติล จริยศาสตร์คุณธรรม และจริยศาสตร์แห่งความอาทร ซึ่งไม่ได้มีฐานคติแบบนี้) ดังนั้นถึงที่สุดแล้วถ้ามนุษย์ตกลงกันได้ว่ากฏทางจริยธรรมที่ทุกคนรับได้คืออะไร เช่นจะเอาหลักมหสุขของประโยชน์นิยม หรือเอาหลักการไม่ใช้คนเป็นเครื่องมือของค้านท์ เราอาจคิดได้ว่า AI ในระบบรถไร้คนขับอาจจะทำงานได้ เช่นถ้าเราตั้งโปรแกรมให้รถสามารถคำนวณความเสี่ยงของทุกคนที่อยู่ในสถานการณ์ที่เกิดอุบัติเหตุนั้นได้ รถอาจจะเลือกได้ว่าจะชนใครถ้าต้องชนคนใดคนหนึ่ง เช่นคำนวณจำนวนคนเอามาถ่วงน้ำหนักกับโอกาสที่แต่ละคนจะมีชีวิตรอด อะไรทำนองนี้ หรือถ้าเราตั้งโปรแกรมให้รถไร้คนขับเป็นรถที่ถือจริยศาสตร์แบบค้านท์ รถอาจจะไม่ต้องคำนวณอะไรเลย จะพุ่งชนคนที่อยู่ข้างหน้าเท่านั้น (ถ้าเบี่ยงแล้วไปชนคนอื่นที่มีจำนนวนน้อยกว่า หรือมีโอกาสที่จะบาดเจ็บน้อยกว่า อาจจะเป็นการใช้คนเป็นเครื่องมือได้) เป็นอันว่าถ้าตกลงกันได้ถึงหลักการแรก แซม แฮริส ก็อาจจะมีประเด็น เพราะเขาบอกว่าที่เหลือหลังจากนั้นเป็นเรื่องข้อเท็จจริง (facts) เท่านั้น ซึ่งในจุดนี้วิทยาศาสตร์คิดแทนปรัชญาได้ เราอาจคิดต่อไปจากแฮริสได้ว่า AI ก็อาจจะคิดแทนมนุษย์ได้ ถ้าคำถามทางจริยธรรมคือคำถามที่ว่า การกระทำที่ถูกต้องคืออะไร ที่เหลือก็เป็นการคำนวณหาความสุขมากที่สุดบ้าง หรือเป็นการเอาหลักการมาใช้กับทุกสถานการณ์แบบตรงไปตรงมาและไม่มีข้อยกเว้นบ้าง ไม่ไกลเกินจินตนาการที่เราจะคิดว่า AI ในรถไร้คนขับอาจจะตัดสินทางจริยธรรมแทนมนุษย์ได้

อย่างไรก็ดี ในบทความเรื่อง "The Trolley Problem Will Tell You Nothing Useful About Morality" (Rennix & Robinson, 2017) เรนนิกซ์และโรบินสันชี้ให้เห็นว่า แม้การทดลองทางความคิดเช่นนี้จะมีประโยชน์ในการสอนหลักการของประโยชน์นิยม คือหลักมหสุข และหลักการของค้านท์ คือการไม่ใช้คนเป็นเครื่องมือ แต่ถ้าเราเข้าใจว่าคำถามทางจริยศาสตร์เป็นคำถามแบบเช่นที่ว่านี้ เรากำลังเข้าใจผิด และถ้านักจริยศาสตร์ทำให้วิชาจริยศาสตร์เป็นการคำนวณหาการกระทำที่ถูกต้องเพียงการกระทำเดียว และมองแต่ละการกระทำว่ามีผลกระทบที่มองเห็นล่วงหน้าได้ชัดเจน (จะให้กี่คนตาย เป็นต้น) พวกเขาก็กำลังทำให้วิชาจริยศาสตร์เป็นวิชาแปลกประหลาดที่ไม่เกี่ยวกับประเด็นทางจริยธรรมในชีวิตจริงของคน เรนนิกซ์และโรบินสันเสนอว่า ชีวิตจริงของคนนั้น ไม่พบเจอเหตุการณ์หายนะสาหัสที่ต้องเลือกให้คนใดคนหนึ่งตาย ถ้าพบสถานการณ์อย่างนั้นจริง คำถามว่าจะเลือกให้ใครตายเป็นคำถามที่ไม่มีความสำคัญเท่ากับคำถามที่ว่า บริบทของโลก เช่นระบบเศรษฐกิจ สังคม การเมือง ใดที่ทำให้เกิดสถานการณ์นั้นขึ้นตั้งแต่แรก เช่น ทำไมคนงานซ่อมรางรถไฟไม่ได้รับสวัสดิการในแง่ของการป้องกันรักษาความปลอดภัยในการทำงาน ทำไมบริษัทรถไฟไม่ถูกตรวจสอบโดยรัฐ เราควรมีความรู้สึกตอบสนองต่อสถานการณ์ ไม่ใช่แกล้งทำเป็นว่ามนุษย์สามารถเป็นผู้ตัดสินที่เป็นกลาง ไร้อารมณ์ จากภายนอก มาตัดสินอย่างไม่เข้าใครออกใครว่าใครควรจะตาย ในบทความที่น่าสนใจนี้ผู้เขียนทั้งสองสื่อนัยว่าความผิดปกติของการทดลองทางความคิดแบบนี้ (และขยายไปถึงความผิดปกติในการเรียนการสอนจริยศาสตร์แบบกระแสหลัก) คือการทำให้เราคิดว่าจริยศาสตร์เป็นการแก้ปัญหาแบบคณิตศาสตร์ คือเป็นการเอาสูตรไปใช้กับสถานการณ์เฉพาะ หรือเป็นการคำนวณผลลัพท์ในแบบที่จะก่อความสุขมากที่สุดหรือก่อความทุกข์น้อยที่สุด ในความคิดของผู้เขียนแล้ว นี่เป็นการไม่เคารพต่อชีวิตคน นักจริยศาสตร์ควรจะแสดงการเคารพต่อชีวิตคนมากกว่านี้ และความผิดปกติประการที่สอง คือเป็นวิธีคิดที่ผู้เขียนเรียกว่าเป็นเสรีนิยมใหม่ (neo-liberal) ในความหมายว่าเป็นการแยกตัวปัจเจกบุคคลออกมาจากสังคมและจากบริบทต่างๆ ทำให้ปัจเจกดูเหมือนเป็นผู้ตัดสินใจในความว่างเปล่า ปราศจากปัจจัยอื่นๆที่จะมามีผลต่อสมการทางจริยศาสตร์นั้น คือปัญหาจะมีเพียงว่า จะให้คนกี่คนตาย คำถามแบบ Trolley Problem ไม่เปิดทางให้ถามว่า ทำไมคนงานถึงได้อยู่ในสภาพการทำงานที่ไม่มีความปลอดภัย

ข้าพเจ้าคิดว่าความเห็นของผู้เขียนน่าสนใจมาก ก่อนอื่นผู้อ่านต้องเข้าใจก่อนว่าผู้เขียนมีความเข้าใจว่าการทดลองทางความคิดคืออะไร และเข้าใจว่านักปรัชญาสร้างการทดลองทางความคิดเพื่อปิดโอกาสต่างๆนานาที่เป็นไปได้ และบีบให้คนฟังต้องเลือกทางหนึ่งในสองทางเท่านั้น เพื่อจะเป็นการทำความคิดของคนฟังให้ชัด และเพื่อสื่อสารประเด็นแก่คนฟังอย่างชัดเจน นักปรัชญาที่ใช้การทดลองทางความคิดไม่ได้คิดว่าชีวิตจริงของคนเราจะไปอยู่ในสถานการณ์ Trolley Problem จริงๆ เรนนิกซ์และโรบินสันก็ไม่ได้คิดว่าการทดลองทางความคิดไม่มีประโยชน์ในการสอนปรัชญา แต่ประเด็นของเขาคือ ปรัชญาจริยศาสตร์ที่เขาเรียกว่ามีลักษณะเป็นเสรีนิยมใหม่ (ในความหมายที่มองคนเป็นปัจเจกบุคคล มีแต่ปัญญาที่ไม่เข้าใครออกใคร เป็นสิ่งที่มาคำนวณความถูกผิดทางจริยธรรม และมองเห็นสถานการณ์ทางจริยธรรมว่าโดดเดี่ยวออกมาจากบริบทอื่นๆ) ไม่ได้ให้ความเข้าใจแท้จริงเกี่ยวกับจริยธรรม (จึงให้ชื่อบทความว่า The Trolley Problem Will Tell You Nothing Useful About Morality) ผู้เขียนทั้งสองเสนอว่าในชีวิตจริงทางจริยธรรม เราต้องดูบริบท ต้องถามคำถามอื่นๆที่นอกกรอบออกไปจากทางเลือกสองทางที่การทดลองทางความคิดกำหนด คำตอบทางจริยธรรม ไม่ใช่การทำสอบแบบปรนัย ที่เราจะเลือกข้อ ก. หรือข้อ ข. แต่ต้องเป็นคำถามปลายเปิด ถามว่า เราจะมีความเห็น ทัศนคติ มีข้อสงสัย มีความคิด ที่เหมาะควรอย่างไรต่อสถานการณ์นั้นๆ เช่นถ้าคนงานต้องทำงานในสภาวะไม่ปลอดภัย แต่ความไม่ปลอดภัยของคนงานไม่อยู่ในสมการของ Trolley Problem เราก็ควรจะถามคำถามนอกกรอบออกไปจากสมการนั้น ผู้เขียนสื่อว่า มันไม่สำคัญด้วยซ้ำไปว่าเราจะชนคนหนึ่งคนหรือห้าคน มันสำคัญที่ว่าเราจะถามอะไร คิดอะไร รู้สึกอะไรหลังจากนั้น ประเด็นของผู้เขียนจึงไม่ใช่เป็นการวิจารณ์ Trolley Problem แต่เป็นการวิจารณ์จริยศาสตร์ที่พวกเขาเรียกว่ามีลักษณะเป็นเสรีนิยมใหม่ โดยใช้ Trolley Problem มาเป็นสื่อในการวิจารณ์

ความคิดของเรนนิกซ์และโรบินสัน ในส่วนที่เกี่ยวกับการตอบนอกกรอบความคิดแบบสมการ หรือความคิดที่ว่าปัญหาทางจริยธรรมแบบยุคภูมิธรรมเป็นปัญหาแบบปรนัย (เลือกข้อใดข้อหนึ่ง) ไม่ใช่เป็นความคิดใหม่ นักคิดอย่างแครอล กิลลิแกน (Carol Gilligan, 1993) ได้กล่าวไว้ในหนังสือมีชื่อเสียงของเธอ In a Different Voice: Psychological Theory and Women's Development ในหนังสือเล่มนี้ กิลลิแกน เสนอว่าการทดลองทางความคิดทางจริยศาสตร์แบบที่บังคับให้ผู้ฟังต้องเลือกทางเลือกอันใดอันหนึ่ง ได้มองข้ามมิติสำคัญของสถานการณ์นั้นๆไปอย่างชัดเจน กิลลิแกนบอกว่าวิธีมองโลกแบบนี้ไม่ใช่เป็นวิธีที่เป็นตัวแทนของการใช้เหตุผลของมนุษย์ แต่เป็นตัวแทนของการใช้เหตุผลแบบผู้ชาย เธอเสนอว่าผู้หญิงไม่ได้มองปัญหาทางจริยธรรมว่าเป็นปัญหาทางเลือกแบบปรนัย แต่ผู้หญิงจะใส่ใจกับบริบทของเรื่องราว และอยากแก้ปัญหาที่เกี่ยวกับความสัมพันธ์ในสังคม ผู้หญิงมองว่าการมีคนป่วยต้องตายเพราะไม่มียารักษา ไม่ควรทำให้เราต้องถามว่าจะขโมยยาหรือจะปล่อยให้คนตาย แต่ควรเป็นเหตุให้เราถามว่า ทำไมสังคมจึงปล่อยให้มีคนต้องตายเพราะจน ในแง่นี้เสียงของเรนนิกซ์และโรบินสันก็สะท้อนสิ่งที่กิลลิแกนกล่าวไว้ คือเป็นความพยายามออกจากคำถามแบบเลือกสิ่งนี้หรือสิ่งนั้น มาเป็นคำถามที่ถามหาความเข้าใจว่าสถานการณ์นี้เกิดมาได้อย่างไรตั้งแต่แรก มีมิติใดในสถานการณ์นี้ที่เราควรให้ความสนใจ มีคนไหนที่ลำบากที่เราจะต้องปกป้องดูแล เมื่อวางกรอบคำถามเช่นนี้ การคิดว่าถึงที่สุดแล้วเราจะเลือกใคร (เลือกให้รถไฟแล่นไปทางไหน) ไม่ใช่คำถามที่สำคัญที่สุด และไม่ใช่คำถามที่น่าสนใจที่สุดทางจริยธรรม

สิ่งที่เรนนิกซ์และโรบินสันและกิลลิแกนพยายามชี้ให้เห็นคือความต่างระหว่างการมองจริยศาสตร์เป็นการแก้ปัญหา (solution) ซึ่งสื่อนัยว่าทุกปัญหามีทางแก้ที่ถูกต้องเพียงอย่างเดียว กับการมองจริยศาสตร์เป็นการเข้าใจปัญหา หรือการมองเห็นปัญหาอย่างถูกต้อง หมายความว่ามองเห็นว่าในสถานการณ์นั้นมิติที่สำคัญมีอะไรบ้าง เช่นถ้ามีคนจนไม่สามารถหายามารักษาตัวได้ การจะขโมยหรือไม่ขโมยยาไม่ใช่คำถามเดียว และไม่ใช่คำถามที่สำคัญที่สุด เพราะมันเป็นสถานการณ์ที่ไม่มีสิ่งที่เรียกว่าคำตอบ "ถูกต้อง" (คือไม่ว่าจะทำอะไรก็ต้องผิดต่อคนบางคนทั้งสองทาง) เช่นเดียวกับปัญหา Trolley Problem ที่ไม่มีคำตอบที่ "ถูกต้อง" (ไม่ว่าจะชนใครก็ผิดทั้งสองทาง) ปัญหาทางจริยธรรมหลายครั้งแก้ไม่ได้ด้วยสมการทางคณิตศาสตร์ (อันที่จริงแล้วแทบจะไม่มีเลยที่จะแก้ได้ด้วยสมการทางคณิตศาสตร์) เราควรถามคำถามอื่นเช่นที่เรนนิกซ์และโรบินสันเสนอ คือถามว่าเรามาอยู่ในสถานการณ์นี้ได้อย่างไร (เป็นคำถามเชิงโครงสร้าง) แต่เราจะไม่เห็นในมิติที่ควรจะเห็น (คือคนซ่อมทางรถไฟน่าสงสาร) ถ้าเราไม่มีความรู้สึก หรือไม่มี "ความรู้สึกนึกคิด" (อริสโตเติลเรียกสิ่งนี้ว่า "ratiocinative desire หรือ desiderative reason" คือ การคิดที่ประกอบไปด้วยความรู้สึก หรือ ความรู้สึกที่รู้คิด)

มาร์ธา นุสบัม (Nussbaum, 1995) แสดงความคิดในแนวนี้ไว้ในบทความ "Poets as Judges: Judicial Rhetoric and the Literary Imagination" เธอได้ยกตัวอย่างสถานการณ์ทางจริยธรรมหลายตัวอย่าง เช่นในสหรัฐช่วงการแบ่งแยกผิว หรือช่วงที่มีการบังคับใช้กฏหมาย Jim Crow's Laws ซึ่งบังคับให้คนผิวขาวและผิวดำแยกกันอยู่ เรียนที่เรียนเดียวกันไม่ได้ นั่งโต๊ะเดียวกันไม่ได้ในร้านอาหาร นุสบัมเล่าว่าผู้พิพากษาศาลสูงสหรัฐในยุคนั้นถามคำถามผิด คือถามว่าคนผิวขาวจะไปโรงเรียนของคนผิวดำได้ไหม คนผิวขาวจะไปนั่งที่ร้านอาหารของคนผิวดำได้ไหม ถ้าต่างฝ่ายต่างข้ามไปในพื้นที่ของกันและกันไม่ได้ นี่ถือเป็นความเท่าเทียมกัน และกฏหมายแบ่งแยกคนผิวขาวและคนผิวดำจึงไม่เป็นการเหยียดผิว และไม่เป็นการกีดกันทางเชื้อชาติ ไม่ผิดรัฐธรรมนูญสหรัฐ (กฏหมายนั้นจึงเรียกว่าเป็นกฏหมาย แบ่งแยกแต่เท่าเทียม (separate but equal)) นุสบัมบอกว่านี่เป็นความล้มเหลวทางจริยธรรมอันเนื่องมาจากการพยายามคิดถึงประเด็นทางจริยธรรมว่าเป็นสมการทางคณิตศาสตร์ และมองว่าในแต่ละสถานการณ์มีตัวแปร มีมิติ ที่พึงให้ความสนใจอยู่อย่างเดียว แต่ในความเป็นจริงในแต่ละสถานการณ์ทางจริยธรรมมีสิ่งที่มีน้ำหนัก มีสิ่งที่ต้องให้ความสนใจ หลายอย่าง หลายมิติ (นี่เป็นเหตุผลที่การทำโจทย์แบบปรนัยจะเป็นความล้มเหลวทางจริยธรรมเสมอ เพราะกำหนดไว้ก่อนว่าสิ่งที่เป็นคุณค่าคืออะไรได้บ้าง) ในกรณีนี้ผู้พิพากษาควรเห็นว่ากฏหมายแบ่งแยกสีผิว เป็นการกีดกันคนผิวดำ เพราะคนผิวดำไม่มีโอกาสที่ดีเท่ากับคนผิวขาว ในการเข้าโรงเรียนดี ได้ทำงานในที่ดี และคนผิวดำมีสถานภาพทางเศรษฐกิจและสังคมต่ำกว่าคนขาวมาก การกีดกันคนขาวและคนดำ ย่อมเป็นการตอกย้ำดูถูกคนผิวดำ ซึ่งเคยเป็นคนชนชั้นทาสมาก่อน นุสบัมกล่าวว่ามิติต่างๆเหล่านี้ไม่ถูกให้ความสนใจ ถ้าเราคิดว่าปัญหาทางจริยธรรมคือการแก้สมการทางคณิตศาสตร์ ปัญหานี้จะเห็นชัดในกรณีการล่วงละเมิดทางเพศ เช่นที่เกิดในโรงงานบริษัท General Motors เมื่อคนงานชายพูดตลกทางเพศกันต่อหน้าวิศวกรผู้หญิง วาดรูปที่ส่อไปในทางเพศที่ลอกเกอร์ของเธอ ไม่มีกฏใดจะบอกได้ว่าผู้ชายพูดคำว่าอะไร หรือวาดรูปอะไร แล้วจะผิดกฏหมายเพราะเป็นการละเมิดทางเพศ แต่บรรยากาศทั้งหมดในที่ทำงาน ทำให้คนที่มีความรู้สึกนึกคิด เข้าใจได้ว่าวิศวกรผู้หญิงนี้กำลังถูกละเมิดทางเพศ และเธอกำลังถูกข่มขู่ในที่ทำงาน สิ่งที่นุสบัมพยายามจะชี้ให้เห็นก็คือ ปัญหาทางจริยธรรมไม่ใช่โจทย์แบบปรนัย (ซึ่งมาจากฐานคติว่าการมีจริยธรรมคือการแก้ปัญหาที่ถูกต้อง คือเลือกทางเลือกที่ถูกต้อง) แต่เป็นการมองเห็น การอ่าน การเข้าใจ การบรรยาย สถานการณ์อย่างถูกต้อง อริสโตเติลในหนังสือ The Nicomachean ethics (Aristotle, 1998) กล่าวถึงละครโศกนาฏกรรมที่เล่าเรื่องกองทัพกรีกไปตีกรุงทรอย ในระหว่างทางมีทหารถูกงูกัดและล้มลงไปดิ้นรนด้วยความเจ็บปวด โอดิซิอุสซึ่งเป็นผู้บัญชาการทหารและเป็นนักพูดที่เก่งกาจ มีความสามารถพิเศษในการใช้ตรรกะหว่านล้อมคนได้ดีที่สุด ตัดสินใจทิ้งทหารคนนั้นไว้ให้ตายคนเดียวในเกาะๆหนึ่งระหว่างการเดินทัพ ต่อมากองทัพกรีกรู้ว่าทหารที่ถูกทิ้งนั้นมีธนูวิเศษ กองทัพกรีกต้องการธนูวิเศษนั้นเพื่อมารบกับกรุงทรอย โอดิซิอุสจึงใช้ตรรกะหว่านล้อมทหารหนุ่มคนหนึ่งชื่อ นอพโธลิมัส ให้ไปหลอกล่อทหารที่ถูกทิ้ง ทำทีไปผูกมิตร เมื่อพิษกำเริบและทหารที่ถูกงูกัดนั้นลงไปดิ้นทรมาน ก็ให้นอพโธลิมิสลอบขโมยธนูของทหารนั้นและนำมาให้ทัพกรีก นอพโธลิมัสถูกหว่านล้อมจนเห็นด้วยตามเหตุผลของโอดิซิอุสทุกอย่าง ปัญหาคือ เมื่อทหารที่ถูกพิษนั้นลงไปดิ้นอยู่บนพื้นด้วยความเจ็บปวดจริง นอพโธลิมัสทำไม่ได้ เขาเห็นว่าการกระทำที่เขากำลังทำอยู่ น่าเกลียดมาก เขาทนตัวเองที่จะทำไม่ได้ อริสโตเติลบอกว่า เขารู้สึกเจ็บปวดที่จะทำการกระทำเช่นนั้น และอริสโตเติลกล่าวว่า เขารู้สึกได้ถูกต้อง และดังนั้นการกระทำของเขาจึงถูกต้อง นักอริสโตเติลรุ่นหลังให้ความสำคัญกับประเด็นเรื่องบทบาทของความรู้สึกในการตัดสินทางจริยธรรมมาก นุสบัมเป็นต้นแบบการตีความอริสโตเติล ในทางที่คิดว่าอารมณ์กับเหตุผลไม่ได้เป็นต่างหากจากกัน (คนที่คิดว่าอารมณ์กับเหตุผลอยู่กันคนละขั้วจะคิดว่าจริยธรรมอยู่ในเขตแดนของเหตุผล โดยเหตุผลจะทำหน้าที่ปกครองอารมณ์) นุสบัมเรียกอารมณ์ในแบบอริสโตเติลว่าเป็นอารมณ์รู้คิด (cognitive emotion) (Nussbaum, 1990) หมายความว่า ไม่ใช่เป็นอารมณ์โกรธหรืออารมณ์สงสารที่เกิดขึ้นมาตามอำเภอใจในขณะนั้น หรือเกิดมาเป็นอาการทางจิตวิทยาของบุคคลเฉยๆ แต่อารมณ์เกิดขึ้นมาเพราะความเชื่ออะไรบางอย่าง (นอพโธลิมัส ในความคิดของอริสโตเติล มีความเจ็บปวด เพราะเขามองเห็นสถานการณ์นั้นว่าเป็นความน่าเกลียด เขากลายเป็นคนน่าเกลียดและทนตัวเองไม่ได้ แปลว่าเขามีภาพว่าคนที่มีเกียรติยศและมีความสง่างามนั้นควรจะเป็นอย่างไร ในที่นี้อารมณ์ของเขาจึงสื่อความรู้ทางจริยธรรมที่ตัวเขาเองถูกหล่อหลอมมา) แม้อริสโตเติลจะไม่ได้วิเคราะห์รายละเอียดของกรณีนอพโธลีมัสไว้ แต่เราวิเคราะห์ต่อได้ว่าการใช้เหตุผลแบบที่วาดคุณค่าเป็นแบบเดียว (อย่างที่โอดิซีอุสหว่านล้อมไว้ คือโจทย์มีว่ากองทัพจะชนะหรือไม่ชนะสงคราม) ตอนแรกสามารถทำให้นอพโธลีมัสเห็นตามได้ (คือนอพโธลีมัสใช้เหตุผลแล้ว และคล้อยตาม) ต่อเมื่อมาอยู่ในสถานการณ์จริง เมื่อเห็นทหารเจ็บกำลังลงไปดิ้นกับพื้น นอพโธลีมัสจึงเห็นว่าคุณค่ามีหลายแบบ ความน่าสงสารของทหาร มีน้ำหนักทางจริยธรรม นำไปสู่การเห็นความน่าเกลียดของตัวเขา ซึ่งก็มีน้ำหนักทางจริยธรรมเช่นกัน แม้ว่าการรบชนะจะมีค่า แต่ไม่ใช่คุณค่าเดียวที่มีน้ำหนักสำหรับเขา

3. ปัญญาทางจริยธรรมและการมีจริยศาสตร์หลายมาตรวัด

แบรี่ ชวอร์ทซ์ และเคนเน็ธ ชาร์ป กล่าวถึงปัญญาทางจริยธรรมในหนังสือ Practical Wisdom: The Right Way to Do the Right Thing (Schwartz and Sharpe, 2010). เขากล่าวถึงคนทำความสะอาดโรงพยาบาลชื่อ ลู้ค ผู้ซึ่งทำความสะอาดห้องคนไข้ไปแล้ว แต่ถูกพ่อของคนไข้ดุว่าอย่างเกรี้ยวกราดว่าลู้คยังไม่ได้ทำห้อง ลู้คตัดสินใจทำห้องอีกครั้งโดยไม่โต้เถียง เพราะเขารู้ว่า มิติที่สำคัญมิติหนึ่งของเรื่องนี้คือ คนป่วยที่อยู่ในอาการโคม่า ลู้คมองเห็นว่าพ่อของคนป่วยอยู่ในอาการเครียดมาก ไม่ได้นอนทั้งคืน และเพิ่งออกไปสูบบุหรี่ข้างนอก จึงไม่ทันเห็นลู้คทำความสะอาดห้อง ลู้คเข้าใจว่า เป้าหมายของการเป็นพนักงานของโรงพยาบาลจะในตำแหน่งใดก็ตาม ไม่ใช่แค่การทำตามหน้าที่เพื่อจะได้สามารถใส่เครื่องหมายถูกในแบบฟอร์มประเมินการทำงานเท่านั้น แต่เป้าหมายแท้จริงคือสวัสดิภาพและสุขภาวะของประชาชน เขาตัดสินใจให้อภัยพ่อที่น่าสงสารคนนั้น เพราะเขาเข้าใจว่าเป้าหมายของงาน (อันที่จริงคือเป้าหมายทางจริยธรรมของคนทำงานด้วย) คืออะไร และในทางส่วนตัวที่มีความรู้สึกของความเป็นมนุษย์ ลู้ค นึกได้ จำได้ รู้สึกได้ ว่าความกลัวลูกจะพิการหรือกลัวลูกจะตายมันหนักหนาเพียงไร ลู้คคิดว่าการกระทำที่ถูกต้องคือการยอมเข้าไปทำความสะอาดห้องอีกครั้งโดยไม่ปริปากบ่น การเข้าไปทำความสะอาดห้องเป็นมากกว่าการทำตามหน้าที่แบบกลไก (ซึ่งถ้าเป็นอย่างนั้นก็ไม่ต้องทำอีก เพราะทำไปแล้ว) แต่เป็นการแสดงตัวตนทางจริยธรรมของเขาออกมา การกระทำที่ถูกต้องในกรณีนี้ (และในอีกหลายกรณีตัวอย่างที่ ชวอร์ทซ์ และชาร์ป แสดงให้เห็นในหนังสือนี้) ไม่ได้มาด้วยการคำนวณ และไม่ได้มาด้วยการคิดโดยตรรกะประหนึ่งทำคณิตศาสตร์ แต่ได้มาโดยสิ่งที่อริสโตเติลเรียกว่าปัญญาทางจริยธรรม ซึ่งหมายรวมถึงการมีความรู้สึกที่สมควร ต่อสถานการณ์ตรงหน้า ไม่มีใครบอกได้ล่วงหน้าว่ามาตรวัดจะเป็นอะไร เช่นถ้าพ่อคนไข้เมื่อเครียดแล้วแทนที่จะตะโกนใส่ลู้ค แต่กลับล่วงละเมิดทางเพศลู้คแทน สถานการณ์นี้จะเปลี่ยนไป และลู้คไม่ควรตัดสินใจเหมือนเดิม (คือยอมตามโดยไม่ปริปากบ่น) เพราะนั่นคือการละเลยมิติสำคัญอีกอย่างหนึ่งไปคือการนับถือตัวเอง ลู้คต้องเปรียบเทียบว่าในแต่ละสถานการณ์ มิติใดในสถานการณ์มีน้ำหนักเร่งด่วนที่สุด ไม่มีใครบอกได้ก่อนว่ามีกฏ มีหลักการแรก มีคุณค่าพื้นฐานใด จะสำคัญกว่าอันอื่นเสมอไปในทุกสถานการณ์ ในแง่นี้แล้วลู้คก็พูดภาษาเดียวกับที่กิลลิแกนเรียกว่าเป็น "เสียงที่แตกต่าง" (Different Voice) ไปจากการคำนวณในจริยศาสตร์กระแสหลัก และไม่ได้คิดแบบบนลงล่าง (Top Down) เหมือนที่เบรนนิกซ์ และ โรบินสัน กล่าวถึงในบทความเรื่อง Trolley Problem แต่เป็นการคิดแบบที่เห็นความสำคัญของสิ่งที่แตกต่างหลากหลาย และเห็นว่าสถานการณ์หนึ่งๆเกิดขึ้นในบริบท มีเรื่องราวอยู่เบื้องหลัง เราไม่อาจตัดสินได้ถ้าเรายังไม่เข้าใจบริบทของสถานการณ์นั้นๆ การตัดสินทางจริยธรรมจึงไม่ใช่เป็นการตัดสินแบบสากล แต่เป็นการตัดสินแบบที่ให้ความสำคัญกับสิ่งเฉพาะ ในบทความ "Discernment of Perception" นุสบัมกล่าวว่านี่เป็นลักษณะที่ทำให้จริยศาสตร์แบบอริสโตเติลแตกต่างจากจริยศาสตร์แบบกระแสหลักอย่างชัดเจน เพราะอริสโตเติลไม่คำนวณ (calculate) แต่พินิจพิจารณาว่าจะพึงมองสถานการณ์นั้นๆอย่างไร (discern)

นัยของการคำนวณ กับการพินิจพิจารณามีประเด็นสำคัญคือ การคำนวณคือการสื่อนัยว่ามีมาตรเดียวสำหรับวัดคุณค่าของทุกอย่าง เช่นความสุข ภารกิจที่เหลือก็คือคำนวณว่าการกระทำอะไรนำไปสู่ความสุขมากที่สุด ส่วนการพินิจพิจารณาว่าจะพึงมองสถานการณ์นั้นๆอย่างไร มีนัยคือมันเป็นเรื่องยาก ไม่เบ็ดเสร็จเด็ดขาด อาจมองได้หลายแง่หลายมุม และนัยอีกอย่างหนึ่งคือ ที่มันยากและไม่เด็ดขาด ก็เป็นเพราะคุณค่าทางจริยธรรมไม่ได้มีคุณค่าเดียว แต่ในทุกสถานการณ์มีหลายคุณค่ามาแข่งกันเรียกความสนใจของเรา การตีไก่ ถูกมองว่าเป็น "กีฬา" และมีคุณค่าเพราะเป็นศิลปะ มีความงาม เป็นสิ่งที่สืบสานวิญญาณของความเป็นไทย แต่ในสถานการณ์ที่เห็นไก่เจ็บมาก หนีไปไหนไม่ได้เพราะถูกขังไว้ในวงล้อมให้จิกกัดกัน มิติที่เราควรเห็นก่อนคือความน่าสงสาร ถ้าเราไม่สงสาร และสงสารจนกระทั่งไม่เห็นว่ามันเป็นศิลปะ เราไม่อาจเรียกตนเองว่ามีความเมตตาได้ ดังนั้นถ้าความเมตตาเป็นคุณค่าของมนุษย์ เราก็ขาดคุณค่าของมนุษย์ในส่วนนั้น ความซับซ้อนของจริยศาสตร์ในแนวนี้ คือเราไม่อาจตั้งโจทย์แบบ Trolley Problem ได้ เพราะการตอบไปข้างใดข้างหนึ่ง ไม่น่าสนใจเท่ากับว่าเราเห็นอะไร เรามีความรู้สึกอย่างไร และเมื่อชนใครคนใดคนหนึ่งแล้ว เรามีความรู้สึกหลงเหลือ (residue emotion) ที่สมควรหรือไม่ (ความรู้สึกผิด ความสงสาร ความต้องการจะช่วยเหลือเยียวยา)

4. ปัญญาทางจริยธรรมเป็นสมบัติเฉพาะสำหรับมนุษย์

ในจริยศาสตร์สายอริสโตเติล AI ไม่มีทางจะตัดสินทางจริยธรรมแบบมนุษย์ (แบบที่มนุษย์พึงจะทำ) ได้ เป็นเพราะจริยศาสตร์แบบอริสโตเติลไม่ได้วางโจทย์ให้เป็นการแก้ปัญหาแบบปรนัย หรือเป็นการแก้สมการทางคณิตศาสตร์ นักออกแบบเทคโนโลยีเชื่อว่าอีกไม่นานท้องถนนจะปลอดภัยขึ้นเพราะทุกคนจะใช้รถไร้คนขับ และรถไร้คนขับจะตัดปัญญาความผิดพลาดของมนุษย์ ซึ่งเป็นสาเหตุหลักของการก่ออุบัติเหตุ นักออกแบบเทคโนโลยีคิดว่ารถไร้คนขับจะทำให้ถนนปลอดภัยขึ้นอย่างมาก และเป็นสิ่งดีสำหรับทุกคน ขอเพียงแต่มนุษย์ตกลงกันได้ว่าจะตั้งโปรแกรมวางคุณค่าให้แก่ AI อย่างไร เช่น ให้รักษาชีวิตคนนั่งเสมอ ให้รักษาชีวิตคนจำนวนมากเสมอ ให้รักษาชีวิตคนที่มาในทางที่ไม่ผิดกฏจราจรเสมอ ถ้า AI ได้รับคำสั่งแบบนี้ ก็จะสามารถคำนวณได้คล่องแคล่วแม่นยำ ยิ่งแม่นยำยิ่งกว่ามนุษย์คำนวณเอง เพราะมนุษย์นั้นเมื่ออยู่ในสถานการณ์คับขันก็จะตื่นตกใจ ไม่อาจมีความเป็นวัตถุวิสัยได้อย่าง AI (ซึ่งเป็นวัตถุเองเลยทีเดียว) เราจึงต้องมาพิจารณากันว่า AI กับมนุษย์ต่างกันอย่างไร การที่ AlphaGo สามารถเอาชนะแชมป์หมากล้อมชาวจีนได้นั้นชวนให้คิดว่า AI สามารถคิดได้ดีกว่ามนุษย์ ปราการด่านสุดท้ายที่มนุษย์คิดว่าตนเองมีความต่างจากโปรแกรมคอมพิวเตอร์คือการคิดสร้างสรรค์ สร้างงานศิลปะ เราคิดว่าถึงแม้คอมพิวเตอร์จะคำนวณได้ดี แต่สร้างงานศิลปะไม่ได้ แต่โปรแกรม AIVA ก็ได้แต่งเพลงคลาสสิกอย่างไพเราะได้แล้ว ความสำเร็จที่น่าตื่นตาตื่นใจของ AIVA มากจากการให้คอมพิวเตอร์ทำตามแบบแผนของเพลงคลาสสิคที่เคยมีมาแล้ว เช่นให้ถอดแบบแผนของเพลงของโมสาร์ต บีโธเฟน บาค เป็นต้น และให้มี reinforcement learning คือให้คอมพิวเตอร์บันทึกว่าผลลัพท์แบบไหนถูกต้องหรือเป็นที่ต้องการ ต่อๆไปคอมพิวเตอร์ก็จะสร้างผลลัพท์แบบนั้นออกมา กล่าวคือคอมพิวเตอร์สามารถเรียนรู้ได้ด้วย ปรับเปลี่ยนตัวเองได้ด้วย

ในทางจริยศาสตร์ ถ้าปีเตอร์ ซิงเงอร์ และนักประโยชน์นิยมจะยอมให้ว่า มนุษย์คิดหลักการแรกออกมา (หลักมหสุข) แต่หลังจากนั้น AI คิดต่อไปเองได้ ก็อาจจะนับได้ว่า AI สามารถคิดตัดสินประเด็นทางจริยธรรมได้ไม่ต่างจากมนุษย์ (อาจมีปัญหาว่าเราจะทำให้ AI เข้าใจได้หรือไม่ว่าอะไรบ้างที่ก่อสุข แต่นั่นก็เป็นปัญหาเชิงปฏิบัติที่เราอาจจินตนาการได้ว่านักออกแบบเทคโนโลยีจะทำได้ในอนาคต เหมือนกับการโปรแกรม AI ให้เล่นหมากล้อมที่มีความเป็นไปได้ในการเดินแต่ละตาเป็นอเนกอนันต์) เราอาจมีรถไร้คนขับสองแบบให้เลือก คือเลือกตามทฤษฎีทางจริยธรรมของผู้ซื้อ คือรถไร้คนขับที่เป็นสายค้านท์ หรือรถไร้คนขับที่เป็นสายประโยชน์นิยม แต่สิ่งหนึ่งที่รถไร้คนขับไม่อาจมีคือ รถไม่อาจมีปัญญาทางจริยธรรมได้ แม้ว่ารถจะตัดสินทางจริยธรรมได้ (คือตัดสินไปทางใดทางหนึ่ง ในสถานการณ์ที่เกี่ยวกับจริยธรรม เช่นจะชนใคร) แต่ไม่อาจมีปัญญาทางจริยธรรมแบบที่เข้าใจในจริยศาสตร์สายอริสโตเติลได้ เพราะประการแรกคือปัญญาทางจริยธรรมต้องเพ่งพินิจคุณค่าหลายอย่างซึ่งไม่ได้อยู่ในมาตรวัดเดียว ไม่ได้มีหน่วยเดียวเพื่อให้คอมพิวเตอร์คำนวณได้ ประการที่สองคือเราจะไม่เห็นว่าคุณค่าที่เกี่ยวข้องในสถานการณ์นั้นๆมีอะไรบ้าง ถ้าเราไม่มีความรู้สึก (คนที่ไม่มีความรู้สึก จะรู้สึกสงสารไก่ชนไม่ได้ รู้สึกถึงความคับแค้นของคนผิวดำไม่ได้ รู้สึกถึงการถูกคุกคามของผู้หญิงไม่ได้) ถ้าเราไม่มีความรู้สึก เราจะเข้าใจและเห็นใจคนอื่นหรือสัตว์อื่นไม่ได้ นี่เป็นสิ่งที่ทำให้เรากล่าวได้อย่างมั่นใจว่า แม้ AIVA จะแต่งเพลงได้เพราะ แม้คอมพิวเตอร์จะทำงานศิลปะในแบบอื่นๆได้อีกด้วย แต่เราไม่อาจมีปัญญาทางจริยธรรมแบบเสมือนจริงได้ ความรู้สึกเข้าใจและเห็นใจนี้ยังสื่อลักษณะอีกประการหนึ่งที่เป็นของมนุษย์แท้ๆ (มนุษย์ในฐานะสัตว์โลก) นั่นคือการที่ชีวิตมีความอ่อนแอเปราะบาง เราเข้าใจความทุกข์ของพ่อที่เฝ้าลูกในอาการโคม่าในโรงพยาบาล เพราะเราเคยมีคนที่รัก มีคนที่รักเรา เราเข้าใจว่าความตายมีความหมายอย่างไร ถ้าจะมีสิ่งหนึ่งที่เป็นของมนุษย์แท้ๆ นั่นก็คือการเป็นสิ่งมีชีวิตที่มีปัญญาทางจริยธรรม

**รายการอ้างอิง**

Aristotle. (1998) (translated by David Ross). *The Nicomachean ethics*. NY: Oxford University Press.

Gilligan, Carol. *In A Different Voice: Psychological Theory And Women's Development*. Massachusetts: Harvard University Press.

Nussbaum, Martha C. (1990). "The Discernment of Perception: An Aristotelian Conception of Private and Public Rationality" in *Love's Knowledge*. Oxford: Oxford University Press.

Nussbaum, Martha C. (1995). "Poets as Judges: Judicial Rhetoric and the Literary Imagination". University of Chicago Law Review: Vol. 62, issue 4.

Rennix, Brianna and Nathan J. Robinson. "The Trolley Problem Will Tell You Nothing Useful About Morality". Current Affairs: November 3, 2017.

Schwartz, Barry and Kenneth Sharpe. (2010). *Practical Wisdom: The Right Way to Do the Right Thing.* NY: Riverhead Books.