

# 人工智能必须考虑风土性因素\*

## ——以“电车难题”为母题

徐 英 瑾

[摘 要] 对于与自动驾驶有关的伦理学规范的讨论，往往以“电车难题”为母题。不过，西方学界处理电车难题的思路，主要就是在义务论思想与功利论思想之间打转，而这种视野上的局限性也影响了西方人工智能专家开发用以处理道德两难的自动推理系统的思路。而在上述处理进路中被普遍忽视的，便是“风土性因素”在紧急避险的过程中所发挥的作用——尽管这些因素的存在，已经得到了大量经验证据的支持。在这里，“风土性因素”可以被理解为在特定文化与地理场域中具有特定特征的司机关于如何进行紧急避险的缄默习惯。从立法角度看，未来可能出现的针对全自动驾驶系统的法律法规，应当在合适的尺度上容纳上述“风土性因素”，以便降低可能的执法成本。而这一诉求自然就会倒逼今天我国的应用伦理学研究者对我国各地区司机所默认的紧急避险原则进行系统的调研与统计。而从技术角度看，要在为自动驾驶技术匹配的人工智能系统中以某种足够灵活的方式实现“风土性”原则，我们就需要对现有的基于符号人工智能与深度学习的主流技术路径进行大胆挑战。

[关键词] 人工智能 风土因素 电车难题 义务论 无知之幕

[中图分类号] B84/B82 - 057

### 一、 导论

随着人工智能产品日益进入人类生活，关于此类产品的运用所能产生的伦理学后效，也日益引发学界的关注。就目前的情况而言，对于人工智能与伦理学之间关系的综合性研究，已经催生了诸如“机器伦理学”(machine ethics)、“计算道德”(computational morality)、“机器人伦理学”(roboethics)等崭新的新学科分支，并随之产生了一个新的学术概念——“人工道德能动者”(artificial moral agent, 简称 AMA)，指那些承担不同程度道德责任的人工智能体。根据不同类型的 AMA 所承担的道德责任的多寡的不同，AMA 又可以被区分为三个层次：(cf. Cervantes et al., p. 503)

(甲) “内隐道德能动者”(implicit ethical agent)。这是指那些通过硬件层面上的物理构造来实现

---

\* 本文系中国社会科学基金项目“对于通用人工智能与特定文化风土之间关系的哲学研究”(编号 22BZX031)、国家自然科学基金项目“探索研究 AI 伦理对科研环境的影响”(编号 L2124040) 的阶段性成果。

特定功能,以便保护使用者安全性的产品。严格的说,这类产品尚且在人工智能研究的范围之外,因为一辆没安装任何计算机软件、但安全性却颇佳的传统汽车也可以被归类为“内隐道德能动者”。

(乙)“外显道德能动者”(explicit ethical agent)。这是指对于特定的道德规则进行软件编制后的产物。不过,此类编程一旦完成,相关产品只能机械地按照既有的指令行事,无法根据自己对于不同境况的判断而自主决定如何对规则加以变通。

(丙)“完全道德能动者”(full ethical agent)。在这个层面上的AMA可以对自己的行为具有“元认知”,并能够根据环境的变化对既定的伦理规则的执行方式加以变通。

很明显,虽然就目前的人工智能发展现状而言,“完全道德能动者”尚且不太可能立即出现,但这一概念所能引发的哲学兴趣或许也最为浓厚,因为对于这个层面上的AMA的研究,实际上已经牵涉到了对新兴认知科学或人工智能科学与传统伦理学之间关系的整体性考察。此外,对于“外显道德能动者”的研究,也需要动用不少伦理学资源,因为那些有待在编程活动中予以算法化的伦理规范的内容,本身就可能引发持有不同规范伦理学立场的哲学家之间的争鸣。

然而,目前在上述两个层面上展开的主流AMA研究,主要都是基于对于西方主流的伦理学理论的人工智能重构,却相对忽视道德心理学研究的资源。而在规范伦理学的范围内,对AMA的研究产生直接影响的也主要是义务论与后果论这两种规范伦理学立场。在笔者看来,上述这些理论建构都过于抽象,并因此缺乏对于伦理学判断中的“风土因素”的关照。这里所说的“风土因素”,指的是特定伦理主体在进行特定伦理判断时,其所处的特定的人文-自然环境对于其判断进程与结果的影响。相比较而言,西方最新道德心理学的研究却对上述风土因素颇有关注,并因此提出了对于西方传统伦理学框架的挑战——然而,上述研究成果却在目下主流的AMA研究的视域之外。

由于AMA的研究往往会被具体化为对于与自动驾驶技术相关的人工智能程序的研究,而与自动驾驶有关的伦理学辩论又往往以“电车难题”为母题,本文的讨论将自然始自于对于传统伦理学框架中的“电车难题”的重估。

## 二、揭开“电车难题”之“无知之幕”所掩盖的风土性因素

笔者认为,典型的伦理学困境在日常生活中的重要性的凸显,往往以社会资源在特定时空中的窘迫性为必要条件。其相关判断理由如下:伦理学的基本任务乃是调节人与人之间的关系,而使得人与人之间关系变得紧张的一个核心因素,便是现有社会资源不够社会成员分配之用。举个例子,“忠孝何以两全?”之所以能够成为一个伦理学困境,乃是因为特定的伦理主体得以为国家尽忠的资源支出预算与得以为家长尽孝的资源支出预算彼此冲突——因此,相关当事人也就难以做到在两个不同的规范性维度上同时投入足够质量的生命资源。也正因为这一点,选择、取舍与排序乃是所有伦理学困境所无法摆脱的宿命。需要注意的是,虽然资源分配问题并没有在以康德为代表的义务论伦理学家的工作中得到明晰化的表达,却在以边沁与穆勒为代表的功利论理路中得到了高度的重视(功利论伦理学的核心就是:如何合理分配社会资源,以便使得最大数量的社会成员能够获得最大程度的福祉——这一工作宗旨显然预设了使所有社会成员均得到完全满足的丰沛社会资源是不存在的)。另外,值得玩味的是,即使是作为康德哲学后人的罗尔斯,在其理论构建中也非常重视“分配正义”这一话题,这说明罗尔斯的研究也预设了可供分配的社会资源的相对匮乏性(因为只有资源相对匮乏的前提下,“分配正义”才足以构成一个具有挑战性的问题)。

伦理主体手头资源的逼仄性,集中地体现在著名的伦理学问题“电车难题”(Trolley Problem)

中。(cf. Foot) 这个难题说的是: 假设一位有轨电车司机正观察到前面的轨道上被绑了五个人, 而且因为电车的惯性原因, 他已经来不及拉制动闸将车停住。唯一救下这五个人生命的机会就是将电车转到旁边一条车道上去——但是他同时却发现, 那条轨道上也被绑了一个人。因此, 他就必须在让五人死与杀死一人之间做出选择。很明显, 这个困境在物理上就排除了司机能够救下所有人的可能性。换言之, 该困境之所以成为困境, 就因为故事的脚本本身就预设了司机手头得以解决问题的资源是有限的——他不能像超人那样, 用超能力让火车在一秒内停止, 或用意念立即为所有绑在轨道上的受害者解除绳索, 等等。因此, 如果我们要将“救下所有人”视为“电车难题”的理想解的话, 那么除非我们无视该思想实验的原始限制条件, 否则我们就必须承认: 这个理想解是不存在的。

正因为不可能存在针对电车难题的理想解, 所以, 基于普遍人道主义的康德式的义务论思想恐怕就很难为解决该难题提供在实践上有用的思路(很明显, 在资源逼仄的前提下, 以任何人的生命为目的的普遍人道主义肯定无法被施行的)。与之相比较, 后果论思想则至少能够通过对于不同选择所导致的后效的计算, 而得到一个更容易被施行的解决方案。比如, 按照后果论的观点, 既然死一个人就要比死五个来得更“合算”一点, 司机就应当选择将车转到旁边的轨道上去。

不过, 电车难题在文献史上之所以被人谈起, 恰恰就是为了针对后果论的处理方案的薄弱之处。一般的见解认为, 如果司机主动将车转向, 那么他就等于是主动谋杀了在旁边轨道上的那一个受害者——而如果他什么也不做, 虽然这样他肯定会害死五个人, 但这些人的死亡毕竟不是他的主动行为的产物——因为电车本来就是向着原来的方向行驶的。在一个更夸张的电车难题的版本中, 对于“以一命换五命”这一做法的质疑可谓达到了高潮: 假设伦理主体不是那位司机, 而是一位在天桥上的看客, 而他恰好发现电车马上就要压过铁轨上的五个人。这时候他还发现自己身边另外一位看客身材非常魁梧, 足以成为电车的“制动器”。于是, 他就将这位看客推下了天桥, 以他的命换了另外五个人的命。(cf. Thomson) 很明显, 在这个版本的电车难题案例中, 更多的人会觉得“以一命换五命”的做法是不正当的, 因为伦理主体主动推人下天桥的做法已近乎是标准的谋杀行为了。

上述批评的确在某种程度上的确揭露了后果论理路的纰漏。但是, 若想以此为据去论证义务论理路的正确性, 则恐怕亦有操之过急之嫌, 因为义务论者并不能通过对于后果论的批评, 自动导出一个符合普遍人道主义原则的针对电车难题的理想解。毋宁说, 上述批评所真正的揭露的, 乃是某种忽略“风土性”因素的电车难题叙事方式所固有的缺陷——而这一缺陷也为对于该问题的义务论视角与后果论视角所分享。

在具体解释“风土性”这个貌似奇怪的术语之前, 我首先想剖析一下使得电车难题的叙事方式得以成立的两个假设性前提:

前提一: “无知之幕”假设。根据假设, 在相关叙事中所有当事人的社会背景都已被虚化。具体而言, 司机是男是女, 是老是少, 有几年工龄, 那些被绑在铁轨上的受害者是否与司机相识, 或者司机是否知道其中某些人的社会地位(他们其中是否有某位天才科学家?), 统统都在故事叙述者的关心范围之外。

前提二: 电车难题的各个版本之间的可通约性假设。我们前面已经看到了, 为了刁难对于电车难题的后果论解法, 学界提出了原始案例的某些衍生版本(如将天桥上的路人推下去的版本), 以使得“以一命换五命”的做法显得更不符合道德直觉。但这个具有“李代桃僵”色彩的做法显然预设了电车难题的不同版本之间是可以彼此通约的——只要它们在内容上都符合“主动作为, 最终杀死一人以救五人”这一形式条件。

然而, 笔者却认为以上两个前提都是不合理的。但由于对于第一个前提的批评就蕴含了对于第二

个前提的批评，所以，我们的讨论重点还是将放在第一个前提上。不得不承认，在某些情况下，“无知之幕”的确是部分存在的。假设在一个雨夜，能见度不算是太好，司机自然无法看清楚绑在铁轨上的人是谁，而只能抽象地知道前方轨道上有五个人，右边轨道上有一个人。不过，即使如此，所谓的“无知之幕”也不可能遮蔽下述信息：至少司机本人应当知道他所从属的电车公司对于此类问题的处理条例，并清楚地了解到：根据这些条例，怎样处理当下的问题才能尽量减少自己的行政责任。在这个问题上，司机并不是以一个原子式的伦理主体的形式而存在的，而是作为一个庞大的组织架构的一分子而存在的。至于对于作为本案例的解读人的我们来说，即使我们无法从该故事脚本里获取任何关于司机所在的有轨电车公司的确切情报，至少我们也模糊地知道：他必然属于某个公司，并处在该公司的相关制度的约束下——故此，他的最终选择其实只是某种组织行为的“个体化表现而已”。而使得这些知识得以浮现的背景知识，则关涉到了我们对于现代有轨电车制度的整体性理解方式（比如，几乎没有一个具有基本社会常识的人会认为仅仅一个人就可以运营一家电车公司，等等）。上述这些稀薄的知识，实际上已破坏了刚才所说的“无知之幕”——换言之，只有预先获得一些知识，司机才可能获得驾驶电车的资格。由此我们不难联想到，在不同国家与不同历史时期，电车司机都会其所从属的社会组织那里得到不同的操作规范手册，以指导其处理各种各样的紧急避险问题。而在不同文化的影响下，电车驾驶规范也会衍生出不同的优先性次序以便指导当事人更有效地分配有限的机会资源。从这个角度看，与特定器械的操作相互结合的规范，就不仅仅是一种纯粹精神的存在了，而具有一种准物质性的表达方式（譬如，路灯的设置就既具有伦理性，又具有毫无争议的物质性）。对于这种被半物质化的伦理规范，有一个专门的哲学术语称呼之，此即“风土”。“风土”是一个在中国与日本都通用的文化－哲学概念，在当代哲学中基于此概念开发出一整套伦理学体系的是日本哲学家和辻哲郎（T. Watsuji）。（cf. Watsuji）在他的语境中，“风土”是指影响人类行为模式的外部人文－自然环境的总和。与近代孟德斯鸠的“地理决定论”不同的是，和辻哲郎所说的“风土”既有地理与气象的因素，又有文化的因素，同时也包含了大量的人工构造物（如交通系统、邮政系统，甚至还包括饮食与穿着习惯，等等）。很明显，在电车难题中，这里所说的“风土性知识”，具体就是指特定的有轨电车公司制度对于特定的紧急避险场景所制定的应对预案，或者是该社会组织对于此类场景的集体应对习惯。而这一风土性知识的存在，自然就使得电车难题中的“伦理主体”的聚焦点产生偏移，即从涉事司机偏移到了司机所在的社会建制之上。这种偏移自然也导致了相关的行为责任的转移：假设该公司的组织架构具有一种很强的后果论/功利主义预设，那么涉事司机即使做出“以一命换五命”的选择；即使这种选择会让义务论者感到非常不适，涉事司机本人对于相关行为的伦理与司法责任也会被最小化。

有人或许会反驳说，就算我们将伦理主体转移到司机所在的组织架构上，原始版本的电车难题似乎也没有被消除，因为作为特定组织架构的决策者，相关决策人士依然需要在无知之幕下直面电车难题的拷问。但依笔者浅见，即使如此，决策者所具有的最起码的风土性知识，依然会冲破这里所说的“无知之幕”。譬如，至少决策者知道：对于紧急避险事宜的处置权应当控制在涉事司机的手里，而与之不相关的人士是无权干涉此事的。换言之，作为相关社会组织的“外部人”，无关人士自然也处在公司组织文化所处的“风土”之外。与之相对应，在紧急避险的具体场景中，处在原始场景中诸因果节点之外的某个人，当然更是处在该事件本身所具有的微观“风土”范围之外了。全社会大多数成员对于这一点的默认，也就解释了为何在电车难题的某些变种版本——特别是将体重超重者从天桥上推下的版本——中，天桥上的旁观者的紧急避险行为的正当性会遭到普遍的质疑。具体而言，在该版本中，风土性原则得到了两次违背：第一次，行动涉及的主体（天桥上的看客）本来就是处在

电车公司的组织架构所覆盖的“风土”范围之外，因此他本无权越俎代庖；第二次，行动所涉及的客体（天桥上的那个体重超重者）也是处在紧急避险的事态范围所覆盖的“风土”范围之外的，因此他本不应该被牵涉其中，而且也没有任何一个别人有权利将其主动地牵涉其中。

上述对于风土性原则在处理电车难题的过程中所起到的作用的分析，尚且没有涉及对于涉事无辜群众（即伦理客体）所具有的社会组织关系（即“风土关系”）的讨论。但在下面这个版本的电车难题变种之中（cf. Dehghani et al. pp. 1280 – 1286），对于上述要素的讨论就会变得不可或缺了。

假设一支联合国维和部队在某个处于无政府状态的国家执行向难民分发粮食的任务。但因为难民的数量超出了预期，所以，粮食的总量暂且不足，只够 100 人吃。若将粮食运输到 A 村，虽能救下 A 村的 100 人，但是这样却就会导致 B 村的 200 人饿死。在可供分配的物质资源有限的情况下，在物理上就不存在救下所有人的可能性。

该案例与原始版本的电车难题有一些区别。第一，伦理主体对伦理客体的风土结构有大致的了解，即两群有待援助的群众分别属于 A 与 B 这两个不同的地方文化组织——与之相较，在电车难题中，司机对于有待援助的群众的社会背景是基本无知的。第二，粮食本身具有几乎无数多种的分配方案（比如，将 20% 的粮食分给 A 村，80% 的粮食分给 B 村，或将粮食对半分，等等）——与之相较，在电车难题中，司机的处置方案只有“任车前行”与“将车转向”这两种。这两项条件的变化不但使得我们对于“分发粮食难题”的处理手段有了更大空间，甚至使得原始的电车难题的解决目标都发生了嬗变。不难想见，在待救群众自身的风土结构暧昧不明的情况下，每个待救群众就只能被视为一个个被遮蔽了所有的社会关系的孤立原子。在这种情况下，只有义务论或者后果论的价值目标才有机会得到浮现（按照义务论，救援任务是否完成取决于是否所有人都得救；按照后果论，救援任务是否完成取决于是否最大数量的群众已经得救）。但在待救群众自身的风土结构得以初步展现的情况下，“保存尽量多数量的文化”这一目标就有机会被凸显为一个更值得考量的伦理目标。按照这一伦理目标，维和部队所应当做的就是将粮食在两个村落之间进行分配，并以不得不饿死每个村落中的一部分群众为代价，使两个群落的文化都能够得以保留。需要注意的是，基于以下两项考虑，这种处理方案并不是后果论方案的某种变种：（甲）按照后果论的观点，救下 A 村的 100 人与救下 A 村与 B 村的各自 50 人是没有分别的（因为两种方案所导致的死亡人数或者得救人数彼此无甚差别），但是根据风土伦理学的观点，二者之间的区别是很大的（前一种做法只能保持一种文化风土，后者却能保持两种文化风土）；（乙）风土伦理学允许维和部队的官兵根据自身的风土属性，去优先救援与之更具文化亲缘性的难民（即使这种对于救援优先性的分配方案会导致被救难民之总体数量的下降），而典型的后果论的分配方案却是不允许这么做的。

对于“无知之幕假设”的批评到此为止。很明显，该假设大大简化了真实的伦理主体在进行道德抉择时所具有的复杂的风土环境，并因此错过了真正的伦理决断所具有大量重要细节。同时，典型的电车难题的叙事方式，已经先验地排除了儒家的、社群主义的、黑格尔主义的以及马克思主义的个体观（按照上述这些哲学流派中的任何一种，任何一个个体都无法脱离相关的社会网络而存在），因此基于这种武断之排除的电车难题案例，并不是我们切入伦理学研究的合适出发点。与此同时，上述分析其实也已经涉及了对于“电车难题的各个版本之间可通约性假设”的批评——说得更清楚一点，由于这些不同版本自身就涉及到不同的风土条件，因此，它们并不是彼此可通约的。因此，伦理学家也就无权通过炮制无数个这样的案例变种来质疑某种特定的伦理学立场的合理性，而只能默认：对于电车难题的任何一种新的变种的诊断意见，都不能被自动延展到对于电车难题的原始案例的诊断之上。

本节的分析依然是处在“概念分析”的层面上。在下节中，笔者将为一种既尊重风土性因素，又与紧急避险问题有关的伦理学研究进路提供经验层面上的支持。

### 三、支持基于风土特殊性的伦理学研究进路的经验证据

笔者将此类证据分为三类。

证据一（来自道德心理学与文化人类学的证据）。根据道德心理学家海特（J. Haidt）的观点，电车难题对于生命伤害问题的过度关注，会使得道德规范所具有的多样性遭受忽略。同样是在演化心理学与社会生物学理论的启发下，海特辨识出了五组先天的道德规范内容（对于这些内容的执行，据说都能增强人类的生物学适应性）：（甲）关爱弱者，避免痛苦；（乙）维护公平，阻止欺骗；（丙）维护忠诚，阻止背叛；（丁）尊重权威，反对颠覆权威；（戊）维护圣洁，阻止堕落。（参见海特，第164页）。海特同时指出，人类文化多样性之所以存在，便是因为上述五项道德规范内容的优先性排序方式在不同的文化中呈现出了不同的体现方式。比如，在一个为自由主义文化为基色的文化场域中，（甲）与（乙）的地位就会被凸显，而在一个以传统威权主义为基色的文化场域中，（丙）、（丁）和（戊）的重要性相对凸显。很显然，从海特的上述立论出发，我们就立即可以得到一个推论：并不存在着一种针对电车难题的“跨文化有效”的解答方案。举例来说，在一个强调权威的文化系统中，如果司机要违背公司上层所制定的以后果论为底色的紧急避险原则的话，那么他由于这种违背而在规范（甲）与（乙）方面得到的受益，将无法抵消他因为违背规范（丙）与（丁）而遭遇到的损失。因此，在这种情况下，他便有更强的倾向性去按照后果论的要求去“以一命换五命”，而尽量压制他对于一条生命的损失而产生的心理遗憾。但在一种自由主义理念强盛的文化氛围中，上述抉择则会被视为是不合适的。而在下文所将提到的由麻省理工学院所主持的以“道德机器”为题的全球性网络问卷调查中，上述理论其实是得到了更全面的例证。

证据二（来自问卷调查的证据）。哲学家布尔特（D. Bourget）与查尔莫斯（D. Chalmers）发起调查，得出结论：在于2009年的以电车问题为内容的测试中，有69.9%的调查参与者会倾向于牺牲一命以换五命，8%的则不会采取任何行动，其余24%的人表示有其他观点或无法回答。虽然从表面上看来，后果论的解答方式在西方职业哲学家之中占据了统计学优势（参加上述测试的人员来自西方99个著名哲学系的受过哲学训练的人士，其中英语世界哲学系占89个），但如果我们在西方一般群众中进行同类问卷调查，会发现一般群众偏好后果论解决方案的比例甚至要高得多。（cf. Bourget and Chalmers）譬如，心理学家纳瓦雷特（C. D. Navarette）在密歇根州对147名美国普通被试所做的调研中，就有133名被试选择将车转向，以一命换五命，即选择后果论处理方案的被试比例高达90%左右，高出同一反应者在哲学工作者中约20%。有意思的是，根据同一项研究成果，如若被牺牲的那个人与决策者有血缘关系，那么愿意牺牲那个人以救出更多人的被试的比例就骤降到三分之一左右。（cf. Navarette et al.）在笔者看来，如上数据其实已经从两个方面支持了基于风土的伦理学路数在电车难题研究中的不可获取的地位。首先，由于康德的义务论哲学已经在西方哲学系构成了某种“圈内文化”，此类特殊风土结构已经明显影响了被试对电车难题的判断方式，使得相关被试偏好后果论的解决方案的比例被调低；但在这种局域的风土文化的影响范围之外，偏好于后果论处理方案的被试的比例自然就被抬高了。其次，在被牺牲者自身的社会背景信息被揭露的情况下，决策者自身的血缘意识与风土意识就会立即被激活，并由此倾向于作出一种既不有利于后果论又不有利于义务论的决策方案，即既不选择救出最大多数的群众，也不选择思考如何救出所有人，而只是聚焦于如何救出自己的亲戚。

当然，上述调查问卷毕竟只是在英语世界展开的，其代表性或许还会受到质疑。不过，风土性因素对于伦理决策的影响，最近又在麻省理工学院主持的以“道德机器”为题的全球网络调查中得到了更具说服力的展示。(cf. Awad et al.) 该调研小组通过网络调查的方式，统计了来自 233 个国家或地区的被试对不同版本的车难题的反应，调查涉及的语言有 10 种（英、德、法、西、葡、俄、中、日、韩、阿），被加以统计的决策数量有四千万次。研究者将被试分为三大文化圈：西方文化圈（主要是西方基督教文化国家）、东方文化圈（包括中、日这样的儒家文化国家以及印尼这样的伊斯兰文化国家）与南方文化圈（包括说拉丁语系语言的美洲文化与非洲文化）。研究在调查问卷中故意撤去了原始电车难题中的“无知之幕”，细化了可能被牺牲的群众的相关特征标签（如性别、年龄、社会地位、是否有案底、是否正在违背交通规则，等等）。研究清楚地揭示东方文化圈的被试，明显有待被牺牲的群众的人数数量不敏感，而更敏感于他们是不是遵照交通法行事的公民（这一点可能体现了东方文化圈对于社会规则的高度尊重）；西方文化圈中的被试则更倾向于什么都不做（这一点可能体现了义务论理念在西方文化中扮演更重要的角色）；而南方文化圈的被试在面对电车难题时，更倾向于保护女性（这一现象的解释还有待研究）。

证据三（来自神经科学的证据）。根据格林（J. Greene）等人在核磁共振成像设备的支持下所做的神经科学的研究，在面对电车难题的不同版本的拷问时，被试大脑的双路向决策通路中的不同决策通路会被激活：在面对原始版本的车难题时，被激活的乃是与推理相关的脑区（比如说前额叶）；而在面对需要牵涉到牺牲局外人（如天桥上的胖子）的道德决策场景时候，被激活的乃是与情感有关的脑区（如杏仁核、脑岛，等）。(cf. Greene et al.) 这非常有力地证明了不同版本的车难题之间是不可通约的，而这种不可通约性就为基于风土特异性的伦理学研究路数提供了解释空间。不过需要注意的是，关于如何从哲学角度解读上述神经科学研究成果，笔者的观点与格林内本人相左。根据他的观点，我们在面对某些版本的车难题时所涌现的那种阻止我们将局外人推下天桥的强烈道德直觉，所印证的乃是康德式的道德义务论的普遍有效性。(cf. Greene) 但在笔者看来，格林内在这里已然忽略了康德式义务论思想的下述基本原则：用以进行道德判断的实践理性本身依然是一种理性，而不是道德直觉。因此，无论对于“不能将胖子推下天桥”这一直觉的解释是什么，这一解释绝不可能是基于义务论的。不过在下述问题上我并不否认格林的观点，即我们正在寻找的这种解释完全可能是植根于演化论的。然而，只要我们由此引入了哈密尔顿的“亲属选择理论”（cf. Hamilton），我们就无法不引入基于风土性或血缘性的行为特异化了：A 对 B 产生正向道德情感的强度，取决于二者之间的血缘关系的密切程度，因此道德主体不会对任何一个道德客体给予同样的关怀。

针对笔者在上文中所展现的这三方面证据，有人或许会反驳说：基于这些证据的上述论证似乎已经预设了一种文化相对主义的立场，而极端的文化相对主义不但会将伦理学的规范性研究还原为纯粹的文化与心理描述（或者与之相关的神经科学描述），而且会将使得我们失去对于某些极端反人类行为（如纳粹在奥斯维辛的暴行）的批判力。对于上述批评，笔者的反驳意见大致有如下两点。其一，至少在讨论电车难题的论域内，我们不太可能会碰触到某些极端反人类的选项，因为即使是用以解决该难题的最为残酷的选项（如将胖子推下天桥），也并不包含着对于特定族群与特定个体的意识形态敌意（换言之，胖子之所以被牺牲，并不是因为他被预先判为敌人，而仅仅是因为他在一个倒霉的时刻出现在了在一个倒霉的位置）。而之所以有时候我们不得不做出某些人必须被牺牲的选择，也仅仅是由于用以覆盖所有被救援者的资源匮乏。其二，本文讨论电车难题的初始动机本来就不是纯粹规范性的，而包含着相关的实践应用目的。换言之，如果我们在 AMA 构造的语境中重新激活电车难题的话，我们就必须考虑到此类 AMA 产品究竟能够给人类的生活带来怎样的便利。这就使得对于此类

AMA 设备的研发过程，将不得不参考道德心理学与文化人类学的视角，而不能光在“扶手椅哲学”（armchair philosophy）层面上进行传统的伦理学研究。令人遗憾的是，目前人工智能界对 AMA 的研究，却普遍没有结合本节所讨论的“风土特异性原则”。而这一点又与西方国家在与人工智能相关的法律法规的制定问题上对于该原则的系统性忽略相关。

#### 四、如何在法律层面上为自动驾驶技术与风土性原则的结合制造空间？

众所周知，自动驾驶技术研究与各国既有法律规范之间的关系是异常密切的，任何与当事国法律相互冲突的自动驾驶技术研究成果都会因为政策原因而自动失去潜在的市场。因此，对于 AMA 研究的具体方向来说，各国的相关政策的引导作用显然是不可忽略的。在全球范围内，目前在人工智能伦理的立法方面，前期工作做得比较多的是欧盟国家，特别是德国。但耐人寻味的是，根据德意志联邦共和国“交通与数码基建部”下属的“自动与遥控驾驶的伦理委员会”颁布的最新（2017 版）伦理规则指导报告，自动驾驶车辆设计所遵循的伦理规范依然是带有强烈的康德的义务论色彩的（并适当兼顾后果论视角下的考量），但几乎完全忽略了风土性原则的影响。这主要体现在该文件的第七款与第九款中。

第七款：在被证明是必然会发生的一些有害场景中，抛开所有的技术性的危害不谈，在对于所有的合法权益之间的取舍进行平衡之时，对于人类生命的捍卫具有至高优先性。因此，在技术可行性的范围之内，相关系统的编程方式，将使得其能够为了防止人命遭到损伤而接受动物生命或者财产的损失。（Federal Ministry of Transport and Digital Infrastructure, p. 11）

第七款体现的康德式人道主义思想貌似能够得到普遍的赞同，却明显忽略了在特定文化中，对某些非人类事物所具有的神圣价值的崇敬，可能会超过对个别人类生命的尊重。设想下述场景：一辆自动驾驶的汽车正迎面遇上一个横穿马路的少年，而若汽车要躲避他，就会撞上路边的一座具有重要文化意义的神龛。因此，该自动驾驶系统就必须在“包含少年”与“保护神龛”之间二选一。笔者不能确定在所有文化中大多数群众都会赞同作出“保护少年，牺牲神龛”的抉择，因此，第七款对于人类生命价值的无条件凸显，很可能无法适合所有的文化风土。

第九款：在不可避免的事故场景之中，对于人类个体的特征（如年龄、性别、其身体与精神状态，等等）的辨认都是被严格禁止的。这么做也是被禁止的：在不同的潜在牺牲者之间进行权衡时，为了保护其中的一者而牺牲另一者。至于旨在减少牺牲者数量的编程方式，则必须得到严格的辩护，才能被加以运用。此外，即使要牺牲被卷入机动车事故的某些相关方的利益，那么没被卷入相关事故的非相关方的利益也绝对不能被牺牲。（ibid.）

根据第九款，相关人工智能系统必须对于事故所可能牵涉到的人类个体的背景信息做到“视而不见”。这就等于恢复了原始电车难题中的“无知之幕”，由此使该系统得以压制在“道德机器”测验中所浮现的不同文化体对电车难题的不同决策倾向。但根据前文所述，在东方文化中的驾驶者更倾向于在遇到各种紧急避险的二难推理时，去牺牲明显违反交通规则的潜在牺牲者的利益（由此显示对社会规范违背者的蔑视），而基于上述第七款的机器编程方式却显然不能满足上述文化诉求。

有人或许反驳说：为何东方文化对社会规范违背者的蔑视具有伦理学上的——而不仅仅是文化人类学上的——意义呢？难道横穿马路者的生命就该被牺牲吗？对于这个问题，笔者的回复有如下三点。其一，从元伦理学角度看，没有任何一种针对紧急避险境况的处理方案具有伦理学意义上的普遍

意义，因为正如前文所揭示的，基于普遍人道主义与义务论思想的相关处理方案也仅仅在西方文化中具有较为明显的支持率。其二，与原始版本的电车难题不同，在城市机动车驾驶的语境中，司机往往看得清路上行人的性别、年龄与衣着，也能够从这些信息中大致判断出其社会地位，遑论对其是否违规横穿马路的判断。因此，第九款条目的执行，会使一台自动驾驶驾车的计算机放弃对上述丰富信息的处理，而只能将这些对象虚化为某些抽象的客体。这样做虽然在技术上并非不可行，却显然与某些重要的伦理直觉——甚至是在西方文化圈内占据优势的伦理直觉——相互抵触（譬如，在遇到紧急避险的情景时，很多西方人都觉得需要优先照顾妇女儿童的生命权）。其三，从技术角度看，一台按照东方价值观运作的自动驾驶系统将比一台按照“条款七”运作的自动驾驶系统能够保护更多的生命，理由是：在某个潜在牺牲者必须被牺牲的紧急情况之下，“条款九”对“无知之幕”的强行植入，将使得系统会因为缺乏相关信息而陷入运作迟滞，并因此导致更大的灾难（正如我们在前文的分析中所看到的，按照义务论思想行事的伦理主体，更倾向于在面对电车难题时什么也不做）。相较于此，东方价值的植入则至少会使系统获得相关指示以进行下一步运作。同时，只要被置入的这种价值观能够在当地获得普遍同情，我们就不能认为此类自动驾驶软件是缺乏特定文化共同体内部的伦理辩护的。

既然德国有关部分所制定的自动驾驶软件的指导伦理规范全面忽略了风土性原则的影响，一种富有中国特色的自动驾驶软件编制的伦理规范应当如何避免这一疏失呢？笔者认为，至少从科技伦理研究的角度看，我们首先需要做的工作，是在前文所涉及的麻省理工学院主导的“道德机器”网络测评的基础上，开展具有中国特色的“道德机器”网络测评，以便为相关的规范设定搜集足够可靠的经验证据。需要指出的是，虽然笔者赞同麻省理工学院主持的“道德机器”网络测评的基本学术思路，但是基于如下考量，我们尚且不能照搬其既有学术成果，而必须根据国情，推行由中国学界主导的同类测评：

第一，既有的“道德机器”检测所涉及的被试对象来自全球，文化分类的线条相对来说比较粗，对中国风土文化的聚焦性还不是很强。

第二，中国国土辽阔，不同城市、地区的驾驶条件与民风民俗都有所不同，因此，在聚焦目标为中国驾驶民间文化习惯的前提下，研究者便有足够的研究精力对被试各自的微观风土结构进行细分，由此得到更为精准的数据。不过，为了保护个人隐私，对于被试的个人身份，相关的检测将不予记录。

第三，问卷未必要以文字方式进行（主要指汉语，也包括主要的少数民族语言），也可以制作模拟不同紧急避险状况的人机互动电脑程序，以实时采集反映被试应答方式的数据。此法也可大大减少因被试阅读能力受限而导致的数据流失。

然而，针对自动驾驶技术的立法工作，只有在具体的智能化自动驾驶技术本身日趋成熟的前提下，才会具有真正的紧迫性。那么，从科学哲学的角度看，我们又该如何向着这一目标前进呢？

## 五、针对人类司机在面对不同伦理问题时应对习惯的建模背后的哲学问题

从科学哲学角度看，在一定文化风土情况下所形成的“习惯”，可以被简化为在一定外部条件刺激下行为主体输出特定行为的趋向（disposition），例如，在所遭遇到的潜在牺牲者是孕妇的情况下，优先躲避孕妇的趋向。这样的一种趋向也可以被表征为“特定条件发生后，特定行为产生的后验概率”。不过，从认识论与发展心理学的角度看，特定心理主体所具有“趋向性”却是一个横跨“先天获得”与“后天习得”两端的模糊概念，因为某些“趋向性”的确可能与遗传相关，而一些“趋向性”则与后天学习有关。倘若我们将针对人类而言的“先天获得”与“后天习得”分别理解为人工

智能语境中的“参数预装”与“系统参数在运用过程中的自行修正”，那么针对“趋向性”之本质的提问就会在人工智能的语境中触发与之平行的如下两个哲学问题。

问题一：趋向性是否需要在编程初始就被编程事先锁死，由此不再变化？从认识论角度看，此问题的历史原型便是近代欧洲唯理论与经验论之争，其关键点在于：人类关于外部世界与伦理善恶的知识，究竟是以先天的方式预存在心智中，还是通过后天经验学习习得的？在这里，只要将一般意义上的“知识”替换为“关于人类处理紧急避险问题的相关知识”，我们便得到了下面这个二难推理：如果反映特定趋向性的数据不是被事先锁死的，那么我们花费大量精力所搜集的反映人类司机的此类趋向的数据，又该如何体现其价值呢？难道除了在将这些数据锁死为被设计的系统的先天知识之外，还有对于它们的更合适的利用方法吗？反之，如若这些数据需要被事先锁死，一旦民众自身的驾驶伦理之趋向性也因某些因素（如一地机动车的道路流动状况）的嬗变而发生嬗变，系统该如何做出及时调整呢？难道要重新进行针对人类司机的新行为趋向进行网络调查吗？由此是否会导致已经完成的建模工作的全面失效，并由此导致巨大的浪费？

问题二：对特定紧急避险情景进行类比性延展问题当如何解决？该问题的语言哲学原型是：如果一个符号运用的主体已经在两个语言表征之间建立起某种逻辑联系，他该如何将这种联系延伸到别的语言表征上？譬如，如果一个人已知“龙虾可吃”，那么他究竟该将这一知识延展到何种程度才算合适？为何从“龙虾可吃”中延伸出“小龙虾可吃”是合适的，从“龙虾可吃”中延伸出“蟑螂可吃”却是不合适的？对于这个问题的 AMA 语境中的重新表达方式如下：有鉴于趋向性数据是附着在特定输入类型与输出类型之间的有限的关系种类之上的，因此若此类数据不被延展，将无法应对一些无法被旧数据框架覆盖的新问题。举个例子：如果适用于牧区的驾驶规范要求自动驾驶车辆将藏獒或者牦牛的生命权重调高的话，那么在此类智能汽车遇到藏羚羊的时候，它该如何应对？“藏羚羊”这个表征在多大程度上与“牦牛”构成了类比关系？很显然，如果系统没有能力在旧表征与新表征之间建立起恰当的类比，它就会时常因为受限于既有的知识库而无法应对外部的新输入；但如果系统做出了一种过于粗疏的类比，它就会经常陷入过于鲁莽的运作。

从某种意义上说，上面所提到的第二个问题乃是第一个问题的变种，因为从旧知识中合理拓展出新知识的能力，在本质上便是如何在“先天—后天”之争中寻找某个合理的妥协点的问题。但我们不得不承认，目前主流的人工智能技术并无力帮助我们找到这个妥协点。具体而言，符号进路的人工智能往往采取“将人类特定领域内知识加以公理化”的进路，并因此无法支持系统对于先天知识的自主性修正——而类似的批评也可以被延展到联结主义—深度学习进路的人工智能之上。大致而言，此类技术进路要求程序员搜集关于人类在特定领域中的行为模式的数据，并将这些数据“喂入”系统，由此使得相关的人工神经网络的内部结构能够被“驯化”到足以模拟人类行为模式的地步。且不提此类技术在“透明性”“可解释性”等方面一直备受诟病，从纯技术角度看，此类技术产品一旦被“训练”完成，也很难产生堪比人类行为的行为迁移性，遑论自行寻找“先天知识”与“后天修正”之间合适的妥协点。（参见徐英瑾，2019年）

为了细化上节所呈现的一般批评意见，笔者在表1中罗列了主流的 AMA 研究所提供的一些代表性的技术成果。这些成果都旨在以自动化的方式处理各种类似电车难题的道德两难处境，但在“是否能够足够灵活的应对文化风土的嬗变所提供的挑战”问题上，似乎都不尽如人意：

笔者自己所认可的技术路线是王培先生发明的“非公理化推演系统”，或简称为“纳思系统”。（cf. Wang）在该系统中，先天知识与后天知识之间的界限乃是相对模糊的。譬如，如果在某个局域的文化风土中，“行车需要躲避牦牛”的确是某种被普遍接受的驾驶习惯，那么搜集到相关数据的纳

表 1 主流 AMA 研究成果一览表

系统名称	主要思路	哲学预设	主要问题
MoralDM ( cf. Blass)	在面对道德二难时，如果系统检测到有“神圣价值”已卷入案例，则输出义务论的解决方案；反之则输出后果论的解决方案。	兼顾义务论与后果论	特定的文化对于“神圣价值”的理解可能发生历史嬗变，而该系统无法自动应对这种嬗变。
Jeremy ( cf. Anderson et al. )	系统对任何一个用以解决道德二难的处理方案所导致的所有相关主体的快乐总量进行计算（这将涉及快乐的强度、持续时间与获得快乐的强度）。	预设了后果论，特别是作为其特殊形式的享乐主义功利论。	不同的选择所牵涉到的快乐的量会随着文化风土的嬗变而嬗变，而对这一点，系统设计者没有加以认真的应对。
LIDA ( cf. Wallach et al. )	对紧急避险情景下的心智活动进行计算建模。核心思想是系统从外部环境提取所识别的对象，使之在系统的记忆池中进行竞争，赢家将获得有限的注意力资源。由此，系统也将决定自己的相关操作对象为何。	并没有在后果论与义务论之间明确选边，而是将道德规范问题还原为紧急避险情景下的心智建模问题。	注意力的分配受到文化风土的影响——譬如，一个印度的司机恐怕要比别文化圈内的司机更留意突然出现在马路上的猴子。系统设计者对于此类影响缺乏考虑。

思程序员就会将这条知识输入系统，并赋予该条知识以比较高的知识权重。不过这并不意味着该条知识是不允许被修正的，一旦系统获得了与上一条知识有所冲突的新信息（这一信息可能是系统自己从对于外部环境的观察中获得的，譬如，它自行观察到最近有不少人行车开始不回避牦牛了），那么系统可以通过特定的“纳思修正规则”而自行调低原始知识的知识权重。纳思系统甚至可以自行对“A 需要被回避”这样的知识进行合理拓展，由此产生“B 也需要被回避”这样的新知识——只要系统能够在“纳思类比算法”的支持下检测到在概念 A 与概念 B 之间存在着足够的可类比性。（关于在此无法全面展开的对具体纳思推理规则的算法细节的介绍，参见徐英瑾，2013 年）

## 六、总结

本文的讨论达成了如下结论。第一，从伦理学的角度看，在西方主流的义务论与后果论思想的框架内，对“电车难题”的典型解决方案将很难顾及“风土性因素”在紧急避险的过程中所发挥的作用——尽管这些因素的存在，并已经得到了大量经验证据的支持。第二，从立法角度看，未来可能出现的针对全自动驾驶系统的法律法规，应当在合适的尺度上容纳“风土性因素”，以降低法规在特定风土中的执行成本。这一诉求要求今天我国的应用伦理学研究者对我国各地区司机所默认的紧急避险原则进行系统的调研与统计。第三，从技术角度看，要在为自动驾驶技术匹配的人工智能系统中以某种足够灵活的方式实现“风土性”原则，需要解决下述问题：系统在已将这些原则所覆盖的特定人类习惯视为某种先天知识的情况下，将如何在外部环境嬗变的情况下，自行调整这些知识？最后，由于目前主流的人工智能技术在解决此类问题时都面临着这样那样的困难，因而笔者为深入解决此类问题提供了新的技术建议。总之，要让人工智能系统真正像人类那样进行紧急避险，我们人类自己首先要勇敢地撕去笼罩在电车难题上的“无知之幕”，并对其背后的风土知识产生“元认知”。

## 参考文献

- 海特, 2014年 《正义之心——为什么人们总是坚持“我对你错”》, 舒明月、胡晓旭译, 浙江人民出版社。
- 和辻哲郎, 2018年 《风土》, 陈力卫译, 商务印书馆。
- 徐英瑾, 2013年 《心智、语言和机器——维特根斯坦哲学与人工智能科学的对话》, 人民出版社。
- 2019年 《试论深度学习技术对人类社会持续发展所造成的风险》, 载《当代美国评论》第1期。
- Anderson, M., Anderson, S. L., and Armen, C., 2004, “Towards Machine Ethics”, in *Proceedings of the AOTP’ 04—The AAAI - 04 Workshop on Agent Organizations: Theory and Practice*, <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=f36b82a0ee77c9a87f3010f9d0335270d1a24687>.
- Awad, E., Dsouza S., Kim R., Schulz, Henrich, J., Shariff, A., Bennefon, J.-F. and Rahwan, I., 2018, “The Moral Machine Experiment”, in *Nature* 563 (1 November 2018).
- Blass, J. A., 2016, “Interactive Learning and Analogical Chaining for Moral and Commonsense Reasoning”, in *Proceedings of the AAAI Conference on Artificial Intelligence* 30 (1), <https://doi.org/10.1609/aaai.v30i1.9817>.
- Bourget, D. and Chalmers, D. J., 2014, “What do Philosophers Believe?”, in *Philosophical Studies* 70 (3).
- Cervantes, J. A., López, S., Rodríguez, L.-F., Cervantes, S., Cervantes, F. and Ramos, F., 2020, “Artificial Moral Agents: A Survey of the Current Status”, in *Science and Engineering Ethics* 26.
- Dehghani, M., T., E., Tomai, E., Forbus, K. D., and Klenk, M., 2008, “An Integrated Reasoning Approach to Moral Decision-making”, in *Twenty-Third AAAI Conference on Artificial Intelligence*, <https://users.cs.northwestern.edu/~mde345/papers/AAAI08-MoralDM.pdf>.
- Federal Ministry of Transport and Digital Infrastructure, 2017, “Ethics Commission: Automated and Connected Driving”, [https://bmdv.bund.de/SharedDocs/EN/publications/report-ethics-commission.pdf?\\_\\_blob=publicationFile](https://bmdv.bund.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile).
- Foot, P., 1967, “The Problem of Abortion and the Doctrine of Double Effect”, in *Oxford Review* 5.
- Greene, J. D., 2008, “The Secret Joke of Kant’s Soul”, in W. Sinnott-Armstrong (ed.), *Moral Psychology, Vol. 3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, Massachusetts: MIT Press.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D., 2001, “An fMRI Investigation of Emotional Engagement in Moral Judgment”, in *Science* 293.
- Hamilton, W. D., 1964, “The Genetical Evolution of Social Behaviour”, in *Journal of Theoretical Biology* 7 (1).
- Navarette, C. D., McDonald, M. M., Mott, M. L., and Asher, B., 2011, “Virtual Morality: Emotion and Action in a Simulated Three-dimensional ‘Trolley Problem’”, in *Emotion* 12 (2).
- Thomson, J. J., 1985, “The Trolley Problem”, in *The Yale Law Journal* 94 (6).
- Wallach, W., Franklin, S., and Allen, C., 2010, “A Conceptual and Computational Model of Moral Decision Making in Human and Artificial Agents”, in *Topics in Cognitive Science* 2 (3).
- Wang, P., 2006, *Rigid Flexibility: The Logic of Intelligence*, Netherlands: Springer.
- Watsuji, T., 1996, *Watsuji Tetsurō’s Rinrigaku: Ethics in Japan*, tr. by S. Yamamoto and R. E. Carter, Albany: State University of New York Press.

(作者单位: 复旦大学哲学学院)

责任编辑: 刘未沫

order in harmony with man , the thought about “a humanity that conforms to order” proceeded from body and emotions to human heart and nature , establishing the validity of the order. Confucian metaphysics , with the search for the basis of order as its pivot , has developed two dimensions which are the heart and inner nature , and transcendental ( *tiandao* 天道 ) , and has unified them into the idea of *Dao*. This can be called metaphysics of order.

**Liu Shipei’ s Notion of Equality: An Analysis Based  
on *The Essential Meaning of Chinese Doctrines of Social Contract***

Fan Guangxin

Liu Shipei’ s embrace of anarcho-socialism in 1907 was not an abrupt choice but was prepared by his reading of a Chinese translation of Rousseau’ s *Du Contrat Social*. Whereas Rousseau’ s messages of freedom and rights inspired Liu to advocate building a democratic republic through revolution , his interpretation of Rousseau’ s notion of equality with the aid of the Chinese populist doctrines led him to endorse “complete equality among human beings” step by step. It was this belief that eventually enabled him to dream of a community of all human beings with no property , no government , and no boundaries.

**Artificial Intelligence , Trolley Problem and the Involvement of Cultural-geographical Factors**

Xu Yingjin

The trolley problem has long played a salient role in the studies of the ethical codes applied to automated driving vehicles. The framework underlying the mainstream narrative for formulating the problem is either deontological or consequentialist. However , neither framework can do justice to the cultural-geographical factors involved in the mental procedure for handling moral dilemmas , although solid empirical evidence supports the existence of these factors. This essay argues that the laws and regulations applied to automated driving vehicles must include these cultural-geographical factors to fit the local cultures. Due to this consideration , a comprehensive survey of local human drivers’ attitudes towards varieties of trolley-problem-alike-moral dilemmas should be carried out before enacting relevant laws. Furthermore , AI scientists also need to develop new approaches to include cultural-geographical factors in automated driving softwares.

**On Egalitarian Theories of Distributive Justice**

Yao Dazhi

Equality is an ideal , but egalitarianism is mainly a theory of distributive justice. On the problem of distributive justice , there are some divergences and controversies among contemporary egalitarians: some persons advocate the egalitarian view , other persons hold sufficientarianism , and the rest of them prefer prioritarianism. The egalitarian view regards equality as the principle of distributive justice , and holds that equality is good and inequality is bad. Sufficientarianism regards sufficiency as the principle of distributive justice , which demands to raise the living level of persons from below to above the threshold , but doesn’ t pay any attention to inequalities of persons above the threshold. Prioritarianism holds that benefiting people matters more the worse off these people are , and so the benefits of the worse off have a priority. Contemporary egalitarians prefer prioritarianism to the egalitarian view and sufficientarianism , but prioritarianism has some problems. The analyses of these problems show that we may construct a better egalitarian theory of distributive justice , which is the negative egalitarianism.