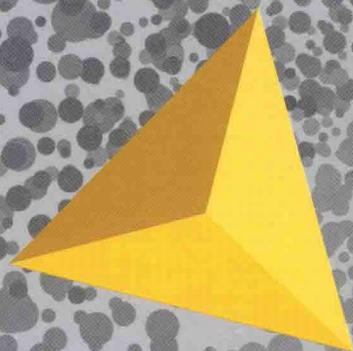




ON WHAT
MATTERS

论
重要
之事



DEREK PARFIT

[英]德里克·帕菲特 著 阮航/葛四友 译

论

ON WHAT
MATTERS

重要
之事

[英] 德里克·帕菲特 著 阮航/葛四友 译

图书在版编目 (CIP) 数据

论重要之事 / (英) 帕菲特著; 阮航, 葛四友译. -- 北京: 北京时代华文书局, 2015.7

书名原文: On What Matters

ISBN 978-7-5699-0185-6

I. ①论… II. ①帕… ②阮… ③葛… III. ①道德—研究 IV. ①B82

中国版本图书馆 CIP 数据核字 (2015) 第 136737 号

北京市版权著作权合同登记号 字: 01-2013-8177

On What Matters

Copyright © Derek Parfit 2011 except:

Introduction © Samuel Scheffler and

Commentaries © Susan Wolf, Allen Wood, Barbara Herman, and T. M. Scanlon 2011

Simplified Chinese copyright © 2015 by Beijing Times-Chinese Press

Published by arrangement with Oxford University Press Inc.

Through Andrew Nurnberg Associates International Ltd.

ALL RIGHTS RESERVED

论重要之事

著 者 | [英] 德里克·帕菲特

译 者 | 阮航 葛四友

出版人 | 田海明 朱智润

责任编辑 | 宋春 陈 媛

装帧设计 | 刘 明

责任印制 | 刘 银

营销推广 | 赵秀彦

出版发行 | 时代出版传媒股份有限公司 <http://www.press-mart.com>

北京时代华文书局 <http://www.bjsdsj.com.cn>

北京市东城区安定门外大街 136 号皇城国际大厦 A 座 8 楼

邮编: 100011 电话: 010-64267955 64267677

印 刷 | 北京京都六环印刷厂 010-89591957

(如发现印装质量问题, 请与印刷厂联系调换)

开 本 | 695mm×995mm 1/16

印 张 | 32

字 数 | 568.5 千字

版 次 | 2015 年 9 月第 1 版 2015 年 9 月第 1 次印刷

书 号 | ISBN 978-7-5699-0185-6

定 价 | 88.00 元

版权所有, 侵权必究

献给汤姆·内格尔和蒂姆·斯坎伦

导 论

塞缪尔·谢弗勒

xix

帕菲特这本书处理了实践哲学的一些最基本问题，其论证详实，且极有原创性。本书分两卷，各有三个部分；其核心章节是第二部分和第三部分，处理的是实质（substantive）道德问题。这些章节源自帕菲特2012年11月在加州大学伯克利分校的三次唐纳讲座。讲座并不包含本书的第一部分和第六部分所处理的问题。第一部分对理由（reasons）和合理性（rationality）^①做了一个扩展讨论，为第二部分和第三个部分的道德主张提供了背景。第六部分探讨的则是元规范（meta-normative）问题。我们在使用规范性（normative）语言来提出有关理由和道德的主张时，就会引出这样的问题。

在本书的第四部分，三位对帕菲特的唐纳讲座做出回应的评论者（斯坎伦 [Thomas Scanlon]、沃尔夫 [Susan Wolf] 和伍德 [Allen Wood]）提供了其评论的修订版。此外，赫尔曼尽管没有参与伯克利事务，但专门为本书写了一组评论。帕菲特在第五部分对这些评论均做了回应。他与评论者之间的交流，主要集中于源自唐纳讲座的那些章节。

帕菲特关于道德的讨论，其目的在于重勘道德哲学的版图。修习道德课

① Rationality、rational和rationally等词的翻译大致是这样的：当涉及人的属性时，一般翻译为理性、理性的和理性地，而当它们涉及其他事物（比如信念或决策）的属性时，一般翻译为合理性、合理的和合理地，但有时候可能涉及的属性所指不明，因此葛译与阮译可能不尽相同，阅读时请注意。（本书所有的页下注均为译者注，本卷的第三部分、附录以及与之相对应的内容提要、尾注与引注为阮航所译，其余的则为葛四友所译。）

程的学生们通常受到这样的教导：后果论者与康德主义者之间存在着根本分歧——前者认为，行动的正当与不当（right and wrong）^①仅仅是其综合后果之函数；后者则通常参考这个或那个版本的绝对命令，从而论证我们具有某些必须履行的义务（obligation）^②，不管这样做是否能够产生后果论眼中的最优结果。尽管人们公认，不管是康德主义观点还是后果论观点都有诸多的变体和改良，但包括康德主义者与后果论者在内的大多数哲学家都认定：两者的差别是深刻且根本的。

在本书的第二、三部分，帕菲特试图破坏这个假定，且试图证实：尽管我们习惯于认为两者的立场是相对立的，但它们之间有着惊人的趋同性。首先，对于康德本人的道德哲学包括绝对命令的各种表述以及其他许多核心观念，他做了一番持久而彻底的考察。尽管康德的道德哲学著作，特别是《道德形而上学奠基》，在道德哲学史上已经得到了最广泛的讨论，但帕菲特对这些文本的处理还是产生了大量新颖的观察与洞见。

帕菲特在序言中交待得很清楚，他对康德的态度很复杂，一言难尽。他把康德描述为“自古希腊以来最伟大的道德哲学家”（235），“短短的四十页却大放光芒，其中给出的富有成效的新观念，比几个世纪以来所有的哲学家提出的还要多”。不过，他马上补上，“不一致性是使康德能取得如此成就的诸多品质之一”。康德的众多评论者都有明确的自身定位——或是作为康德的批评者或是作为康德观点的捍卫者，但帕菲特采取了与他们不同的进路。他把康德的文本当作一座富矿，其中所有的主张、论证和观念，都值得我们就像对待一个卓越同代人的观念那样认真对待。不过，他认为，康德的许多观念需要澄清和修正，有些观念则根本没用了。帕菲特对这些观念、论证和主张做了大范围的考察，让它们经受高水平的检验，这种检验的显著特点在于执着的聚焦和分析的强度。他的主要目的既不是捍卫也不是批评康德，而是要确定，我们可以利用康德的哪些观念来取得道德进步。进步才是帕菲特最终的真实目的。正如帕菲特在解释我们为什么要修正康德的一个表述时所说的：“在学习伟大哲学家的作品之后，我们应尽力取得更多的进步。我们站在巨人的肩膀之上，有能力比他们看得更远”（300）。

① 这里的right和wrong均具有很强的道德色彩，因此译为正当与不当，以区别于一般意义上的正确与错误。

② Duty、obligation在帕菲特这里没有特别的区分，故统一翻译为义务。

帕菲特辨识出康德思想的若干要素——对于这些要素，帕菲特尽管做了某些重大的修正与补充，但仍认为它们特别重要，也乐于认同它们。然而，他诠释这些观念之内容与含义的方式，往往不同于其他重要的评论者。他对绝对命令的“普遍法则公式”（FUL）的处理，也许最清楚地显现了这一点。如帕菲特所见，绝对命令的这一公式一直受到许多严重的反驳；这使得许多本来抱同情态度的评论者都认为，就作为区分正当与不当的指导原则而言，它对我们帮助甚小。许多重要的康德学者都认为，绝对命令的其他表述更为丰富，也更有启发。

相比之下，帕菲特则看到了普遍法则公式的巨大潜力。与流行的诠释意见相左，他坚持我们“能够使得FUL变得可行”，他主张“这个公式以完全康德式的方式修正后，是……相当成功的”（294）。实际上，他走得非常远，以至于认为这个公式经过恰当修正后的版本“也许就是康德自称一直在苦苦寻觅的那个最高道德原则”。帕菲特偏爱普（342）遍法则公式的这一修正版本：“每个人都应当遵循那些每个人都理性地意愿其被普遍接受的那种原则。”这个版本诉诸一种普遍的选择或同意，因此它有资格作为一种形式的“契约论”，帕菲特将其作为“康德式（Kantian）^①契约论公式”。如果如此诠释，康德式立场就可以与契约论的各种当代版本——特别是那些本身受到宽泛意义上的康德主义启发的各种版本——做比较。罗尔斯诉诸在无知之幕后得选的原则就是一例，尽管他把这种装置几乎只用来为社会的基本结构选择正义原则。罗尔斯在《正义论》中曾短暂地持有这样的观点——可以将同样的装置运用于更广泛意义上的道德原则选择——但从未贯彻下去。尽管如此，帕菲特还是对这个观念做了严厉的批评，并断定，就作为一般性的道德论说（account）^②而言，罗尔斯式契约论的前景远不如斯坎伦所发展的契约论。

xxii

按照帕菲特的表述，“斯坎伦的公式”认为“每个人应当遵循无人可以合情理地（reasonably）拒绝的那些原则”。帕菲特认为，至少基于某些诠释，斯坎伦式契约论与康德式契约论是吻合的。因为基于这些诠释最终会发现，每个

① Kant's与Kantian与两个术语的译法分别为“康德的”与“康德式的”（或译为康德主义的），前者强调是属于康德的，后者不仅包括康德的，还包括对康德观点所做的发展与引申，这种处理也同样适用于Scanlon's与Scanlonian等，阅读时请注意。

② Account一般来说比theory弱，但比一般的假说或解释要强，具有一定的结构。为了突出这一点，译者采用了“论说”这一有时候甚至觉得有点别扭的说法，以避免误解。

人都理性地意愿要被普遍接受的那些原则与无人可以合情理地拒绝的那些原则是完全一样的。尽管帕菲特与斯坎伦在这两种契约论趋同的确切程度上存在着分歧，但它们趋同的可能性并不特别让人吃惊。更让人吃惊的是帕菲特对于契约论与后果论之间关系的评价。

正如我已经提到的，康德主义立场与后果论立场之间的对立通常被看作是深刻且根本的。当代的契约论，无论是罗尔斯的还是斯坎伦的，其旨旨在很大程度上都在于阐述替代后果论的某种强有力理论。然而，帕菲特认为，康德式契约论实际上蕴含了某种版本的“规则后果论”，后者认为“每个人都应当遵循这样的原则：其被普遍接受将使事情进展最好”。他强调，每个人都理性地意愿其被普遍接受的那些原则恰恰就是那些“最优的”规则后果论原则。据此，康德式契约论和规则后果论可以组合形成被称为“康德式规则后果论”的观点：“每个人应当遵循最优原则，因为只有它们才是每个人会理性地意愿其成为普遍法则的原则”（411）。尽管就人们应当遵循的原则之主张的内容而言，这种立场是后果论的，但就我们为什么应当遵循这些原则的论说而言，这种立场与其说是后果论的，不如说是康德主义的。我们应该遵循这些原则是因为我们会理性地意愿它们被普遍接受，而不是因为如后果论者所说的，最终重要的就是产生最好的事态。

既然康德式契约论蕴含规则后果论，既然某种版本的康德式契约论与某种版本的斯坎伦式契约论是吻合的，那么三种立场的各个版本就可以组合起来。由此导致的“三重”理论就认为“某个行动是不当的，当且仅当它不容于这样的原则：该原则是最优的，是唯一可普遍地意愿的，且是不可合情理地予以拒绝的”（413）。帕菲特深信，各种可能性的趋同导致：康德主义者、契约论者和后果论者之间有着深刻的分歧是种错误的看法。相反，“这些人是从不同的侧面攀登同一座山峰”（419）。

帕菲特在发展这种核心的论证思路时，极为依赖他有关理由与合理性的实质性（substantive）主张。人们在想要事物或想做事情时持有何种理由，人们的行动在何种条件下是合情理的或合理的，帕菲特考察的各种理论对此都提出了各自的主张。帕菲特对于这些理论的评价，也就很大程度上在于评价这些不同的主张所具有的力度。不过，有关理由与合理性的主张所具有的争议性，一点也不少于有关正当与不当的主张所具有的争议性。帕菲特认识到了这一点，因此在他开始有关道德的章节之前，他对自己有关这些主题的观点做了详尽的阐述与捍卫。

许多哲学家认为，我们的行动理由全部是由我们的欲望来提供的。我们最有理由做这样的事情，它能够最好地满足我们的实际欲望^①，或者满足在理想条件下我们会具有的欲望。帕菲特把这种基于欲望（*desire-based*）的看法称之为“主观理论”。尽管无论是在哲学之内还是之外，这种理论都一直有着深刻的影响，但帕菲特还是相信：它们有着深刻的错误，且他对这种观点的批评是毁灭性的。他主张，这些理论不仅具有大量说不通（*implausible*）^②的含义，而且最终是“建立在浮沙之上”。这些理论蕴含着：我们理由的规范力量来源于我们根本没任何理由具有的欲望。而帕菲特主张，这样的欲望并不能给我们任何理由。由此，最终，基于欲望的观点的真实含义是我们根本没有任何行动理由，由此在最根本的意义上说，没有任何事情是真正地重要的，其意义是指我们没有任何理由在乎任何我们确实在乎的事物。

在拒绝这些“黯淡”的观点之后，帕菲特反过来主张我们应该接受基于价值的（*value-based*）客观理论。按照这种观点，行动理由是由行动所实现或满足的价值（或者用他的话讲，“使得某些事情为了其自身的缘故而值得做”的事实或“使得结果或好或坏”的事实）所提供的。以这种方式理解的话，理由的判断就比合理性（*rationality*）的判断更为根本。因为在帕菲特看来，我们对理由或显见（*apparent*）理由做出回应时，我们就是理性的；而若我们的信念为真，且我们做的事情有好的理由，则我们的行动就是合理的。这与实践理性的许多流行论说形成了对比，例如后者之中有些将合理性确定为期望效用的最大化，有些将实践不理性诠释为某种形式的不一致性。

如斯坎伦文章中的看法，理由优先于合理性的观念也与康德的观点相冲突。对于康德而言，无论是绝对命令的权威还是对其内容的理解都要参考理性的能动性，而不是参考有关“人们所具理由”的某种独立观念。正如斯坎伦所描述的康德主义观点——他称之为“理由的康德式建构观”：“有关理由（确切地说是人们必须看作为理由的因素）的主张必须扎根于人们有关理性能动性的主张，也就是有关人们把自己看作理性行为者相一致而能够采取何种态度的主

① *Desire*这个词在帕菲特这里所指非常广泛，既包括我们与生俱来的各种生理欲望，还包括各种心理上的欲望与愿望，无论是欲望还是愿望不能准确翻译它，这里译为“欲望”是不得已而为之。要注意这种欲望就其一般而言并无好坏之分，但具体的欲望则可能有好坏之分，阅读时请注意。

② *Plausible*、*plausibility*等的译法并未强求统一，一般译为可行的和可行性，有时则译为言之成理的，说得通的等。

张。证成 (justification) ^①永远不能反过来, 即从有关理由的主张来得到有关“合理性要求什么”的主张。

帕菲特像斯坎伦一样拒绝了理由的康德式建构主义, 且如斯坎伦所指出的, 帕菲特表述 (他试图证实会趋同的) 各种道德理论的方式都诉诸了这一观念: 个人能理性地意愿的东西都预设了独立可理解的“有关个人所具理由以及理由相对强度”的观念 (118)。这点把这些理论区别于康德自己的观点, 也区别于某些重要的当代康德主义者 (如科丝嘉) 的观点。帕菲特承认, 他依赖于一种原始且“不可定义”的理由, 随之还承诺存在一种不可还原的规范真理, 这使得他的观点成了某种被科丝嘉称为“独断的理性主义”版本。就此而论, 它不仅受到如科丝嘉这样的康德式建构主义者的抵制, 而且也受到倡导不同元伦理视界 (outlook) (如某种形式的自然主义和非认知主义) 的学者的抵制。

xxv

因此, 帕菲特在第六部分承担着这样的任务: 解释并捍卫自己的规范性观点。他把自己认同的观点称为“非形而上且非自然主义的认知主义”。这种观点诉诸我们对“不可还原的规范真理”具有的某种直觉信念。这种观点并不主张存在某些非时空的实体 (reality) 成分, 在此意义上, 它不是柏拉图式的。此外, 它对直觉的依赖并不是用来表示: 规范事实是通过某种类似于感官感觉的精神能力而理解的。我们也并不把像正当性和合理性这类规范属性的出现看作是规范事实因果影响的结果。相反, 我们对某些事物的规范真理的理解方式就像理解数学或逻辑真理的方式那样。帕菲特实际上主张, 数学推理和逻辑推理本身就包含认识到“我们有理由相信什么”的规范真理, 且要对之做出回应。例如, 我们承认真理p, 且“如果p, 那么q”, 这就给了我们结论性的理由相信q。帕菲特坚持, 正如存在着我们有理由相信何事的真理, 也存在着我们有理由做何事的真理。

当然, 帕菲特也意识到, 许多哲学家并不接受“存在着这种意义上的不可还原的规范真理”。虚无主义者和谬误理论者都认为, 一切规范主张都是假的 (false) ^②。自然主义者认为规范事实可还原为自然事实。非认知主义者则认为, 尽管规范主张在人类生活中很重要, 但它们根本不作为事实性陈述起作

① Justification通译为证成, 而justified、justifiably葛译为初看有点别扭的“有证成的”“有证成地”, 以期读者更好地注意它指综合权衡后更有道理或理由等, 而阮译为求通顺, 有时候译为有正当理由的等。

② True和false当用来指命题时, 通译为“真的与假的”, 其他的则视情况可能译为“对的与错的”或者“正确的与错误的”等。

用。对于这些立场，帕菲特讨论和批评了其中许多有影响的版本，包括布莱克本（Blackburn）、布兰德特（Brandt）、吉巴德（Gibbard）、黑尔（Hare）、麦基（Mackie）和威廉斯（Williams）等人的观点。他主张，这些理论没有一个能够恰当地解释我们思想中的规范性维度；基于所有这些观点，规范性都被证明是种幻觉，消失不见了。实际上帕菲特看来认为：所有这类观点都倾向于虚无主义，且对于承认不可还原的规范真理来说，虚无主义是唯一真正的竞争对手。科丝嘉对规范性“实在论”的康德式反驳也说服不了他。与科丝嘉强调的相反，他声称规范性的根源并不在于意志，反而在于不可还原的（有关我们有理由相信什么、想要什么和干什么）规范真理的存在。

xxvi

正如马上会明朗的，帕菲特讨论理由与规范性的目的，完全不同于讨论实质性道德理论时所追求的目的。在后一种情况下，他的目的是证实公认对立的某些理论实际上是趋同的，因此它们之间的表面分歧就会消散。但是在讨论有关理由和规范性的不同观点时，他并不把竞争理论之间的趋同性纳入议题范围。主张他相反，我们应该接受基于价值的理由理论，拒绝基于欲望的理由理论。同样，我们应该接受他的认知主义理论而拒绝一切形式的自然主义和非认知主义理论。帕菲特明显受困于实质的道德分歧，因为他认为这会威胁到我们对存在道德真理这种事情的确信。这就是为什么他有如此强的动力，要去证实各种相竞争的道德理论具有趋同的可能性。他尽管也受困于元伦理或元规范的分歧，但反应是大不一样的。在此他只是简单地试图决定何种理论是正确的。然而，帕菲特试图证实会趋同的各种理论都预设了他有关理由与规范性的观点，而这种观点显然是有争议的，这就会让我们质疑他所描述的实质层面的那种趋同性是否真是重要的。对那些拒绝基于价值的理由理论的人，或者对那些接受这种或那种形式的自然主义、非认知主义或建构主义的人而言，这种道德共识所依赖的元伦理学观点恰恰是他们所要拒绝的，故而是无法打动他们的。因此，帕菲特面临的一个挑战是要证实他所力证的这种趋同性，其重要性不会被它对有关理由和规范性的（不存在任何趋同性）主张的那种依赖所破坏。帕菲特尽管没有直接应对这个挑战，但确实主张：人们之所以拒绝他所偏爱的那种有关理由和规范性的观点，是因为他们对其观点并没有充分的了解。他表达了这种愿望：一旦相关的误解得以消除，越来越多的哲学家最终会接受这些观点。如果此论不谬，即便理由和规范性的相竞争的理论本身并不趋同，我们也有理由指望，哲学家对这些理论的评价上会有越来越大的趋同性。当然，这个建议本身就很可能有争议的。

xxvii

对于帕菲特精细微妙的论证，我们还能且会提出更多其他的问题。一个问题是帕菲特试图证实会趋同的各种道德观点，在何种程度上是其常见观点的真实版本。四个评论者均从不同方面对这个问题做了讨论。康德式契约论在何种程度上是真正康德式的？我们已经看到，就合理性与理由之间的关系论说而言，这个观点看起来更像是帕菲特的而不是康德的。同样的问题也可以对其他表面趋同的立场提出来。斯坎伦式契约论在何种程度上是斯坎伦自己的观点？帕菲特的规则后果论版本与其他的后果论表述之间的关系又如何？

这是个棘手的问题。正如斯坎伦提到的，帕菲特非常直率地承认，在发展“康德式”立场时，他乐意偏离康德的实际情况，只要他认为这有助于改进它们。如帕菲特所言，“我们在问的是，康德的公式能否有助于我们决定哪个行动是不当的，能否有助于我们解释为什么这些行动是不当的。如果我们能以显然必要的方式修正这些公式，那么我们是在发展某种康德式的理论”（298）。在对斯坎伦的回应中，他同样明确了这个事实：他对于康德式契约论与斯坎伦式契约论的趋同论证“并不适用于斯坎伦书中陈述的观点”（卷二，244），相反是适用于在他看来是这种观点经过修正而加强了版本。

这种不认错的修正主义给帕菲特带来了两种风险。斯坎伦提到的第一种是，他能证实的任何趋同有多么让人吃惊和多么重要，取决于趋同的各种理论与它们的理论原型有多接近。修正的方式越大，理论离原初表述越远，那么这种趋同就越不令人吃惊，也越不重要。第二个风险是修正原初理论让它们更为接近时，我们有可能排除掉了原初理论有价值的成分。

沃尔芙看来很是认为帕菲特的趋同性主张同时有着两类风险。就帕菲特试图调和康德主义、后果论和契约论传统的抱负而言，她写道：“只要上述所引话语是意图表明，这些不同传统所强调的价值能够以这样的方式诠释和排序，从而可以消除它们之间的张力，或者说本着这些传统最伟大的倡导者的精神来接受对所述观念的修正和限定，从而最终与其对手和平共处，那么帕菲特就偏离了他所讨论的哲学家的明确立场，这种偏离方式在我看来既在诠释上是讲不通的，在规范性上也是令人遗憾的”（卷二，32）。沃尔芙的观点是：康德主义、契约论和后果论传统各自有着相异的评价性视野，每种视野都有重要的贡献，但它们彼此之间确实有着真正的张力。这些张力反映了我们道德思想本身之内有着更广泛的张力。她认为，就此而论，无法消除这些张力也是用不着遗憾的。帕菲特试图寻求的那种统一原则将必然是个妥协而不是趋同的问题，并且任何这样的原则必然会遗漏某些重要的价值。沃尔芙特别参考了帕菲特的康

德主义版本来强调最后一点，她认为，这种版本限制了自主性在康德本人的道德哲学中的重要性。

赫尔曼也认为，帕菲特的立场在根本的方式上偏离了康德的立场。然而，尽管沃尔芙怀疑帕菲特寻求的“道德依赖于统一原则”这个观念本身，但赫尔曼赞成康德自己的统一性论说，还认为帕菲特的理论是不同成分的不稳定混合。更具体地说，她认为帕菲特采用了一种“杂交”方法论，纳入了一些康德主义的特征，但尽管如此还是具有强烈的后果论色彩（卷二，81）。尽管帕菲特的意图是保持康德观点中最有说服力的成分，同时避免其中显然不得人心的含义，但赫尔曼还是认为：在康德主义与后果论方法论之间的“不匹配”是深刻的，试图组合它们必然会扭曲康德自己的论说，还会遮蔽这种论说中最有吸引力的成分。在其评论的第一部分，她确定了帕菲特方法论中的几种成分，认为它们具有深刻的后果论特性，然后她对她看到的帕菲特与康德在方法论上的分水岭做了阐明。也许最基本的差别是这一点：帕菲特诉求了各种非道德善来决定我们可以理性地意愿什么，然后依此来决定道德的内容，然而，赫尔曼说道，康德则是试图在一个独立地确立的道德框架内来确定非道德善的地位。在其评论的余下部分，她试图证实，“统一的”康德式进路如果得到恰当发展，那么就有办法纳入康德看来忽略了的某些最重要的道德直觉——诸如有关可允许的谎言的直觉。如果此论不谬，那么我们想要一种混合的道德方法论的动力就被大大削弱了。赫尔曼颇费苦心以这种方式发展统一的康德主义观点，对此帕菲特在其答复中未予直接回应。然而，他拒绝赫尔曼对他的方法论与康德方法论“不匹配”的评价。他强调，赫尔曼所引其明显的后果论方法论中，实际上大多数方面也是康德观点的特征。尽管他对康德的普遍法则公式确实做出了修正，但有些修正是完全本着康德主义观点之精神的，还有一些则是为了避免一目了然的错误。帕菲特认为据之可得出，他与康德在立场之间的差距，远没有赫尔曼所声称的那么广泛和深刻。

像赫尔曼一样，伍德也主张帕菲特的哲学方法论在诸多重要的方面都与康德相背离，但与赫尔曼相比，他更注重帕菲特在方法上的不同之处。伍德认为，帕菲特的方法论起源于西季威克，这种方法给自身设定的目标就是提供一种“科学”的伦理学。这种想法是要把我们常识性的道德意见系统化，有必要时做出修正，其目的在于得到一组确切的原则，由此对于个人在一切可设想的环境下应该如何行动可以用某种算法得到一个确定的道德裁定。伍德认为如康德、边沁和密尔等人本来就是非常不同的哲学家，采取的是完全不同的一种方

法，他本人认为这种方法要优于他归于西季威克和帕菲特的那种方法。这种方法不是始于道德常识，而是始于某种阐明基本价值的根本原则。一般性的道德规则或义务是由此非演绎性地从根本原则派生出来的。这些规则或义务代表了某种努力，其目的在于诠释人类生活条件下的根本价值之蕴含。规则或义务本身容许有例外并且要求诠释：它们在应用于特定情形时需要我们做具体的判断，而无法以确切的规则或原则去规定。因此，如伍德所理解的康德式方法一方面不像西季威克式方法那么重视常识性的道德直觉，另一方面，它认为没有希望达成这样的目标：建立一种能够提供某种算法的“科学的”伦理学，以服务于道德决策。

伍德认为——尽管帕菲特的回应表示他不接受这种诊断——刚描述的这种方法上的差别构成了他与帕菲特在有关康德人性公式的诠释上产生分歧的潜在原因。伍德还认为，这种差别也使得他们对于一种常见类型的哲学论证表现出泾渭分明的态度。这类论证使用我们对某种程式化且有时很复杂的虚拟例子的直觉反应来检验各种道德原则。所有这类例子无论是否涉及实际的列车，伍德都称之为“列车问题”，以揶揄地致意于由富特首次引入哲学文献的这类著名情形。帕菲特频繁使用此类例子来建构他的论证。例如，他对康德主义与规则后果论的趋同就极为依赖这种主张：在一种行动方案会给行动者自己带来负担且唯一的替代方案就是给别人带来负担这样的情境下，个人能够合理地同意什么。帕菲特用一系列虚构例子（大量不同的虚构环境会给不同的人带来负担，其类型不同且程度不一）来阐明和捍卫他的主张。他试图在这些例子中引领我们道德上的直觉反应，以表明：（1）每个人都会理性地意愿后果论的最优原则会被普遍接受，即便这些原则会给自己带来负担；（2）不存在人们会理性地选择普遍接受的任何其他原则。帕菲特显然相信：使用虚构例子有助于澄清在复杂道德选择中的相关问题，并使得我们能够在道德论证中取得进步。相对比而言，伍德认为“列车问题”对于道德哲学一无是处（68）。他的论文的主要部分都在于对这种方式的一种广泛批评：依赖此类问题会把道德哲学家引入歧途。

只要其他人也像伍德一样，对在道德哲学中诉诸虚构例子持保留意见，那么帕菲特大量依赖此种例子就会成为人们抵制其论证的根源。当然，即便是并不像伍德那样完全拒绝此种诉求的那些人，也可能发现，在帕菲特讨论的某些具体例子中，其反应不同于帕菲特，尽管帕菲特已经预见到了许多潜在的分歧并且展示出力图消弭分歧的杰出智谋。然而，帕菲特自己也指出，人们对于某些情形的反应取决于他是接受基于欲望的还是基于价值的理由理论。既然他指望使用我们的反应来支持他有关不同理论趋同的主张，这种变量（我们反应

上的分歧)便反映出一种可能:正如我们对规范判断性质的元伦理分歧一样,有关理由和合理性的分歧会产生威胁,乃至动摇他试图确立的那种道德共识。正如我已经说过的,帕菲特对于这种威胁的回应,不是在相竞争的元伦理学理论、理由和合理性理论本身之间寻求趋同,相反,他主张我们有着决定性的理由,可以拒绝“非形而上且非自然主义”的元伦理学和基于价值的合理性理论之外的一切其他选项。他把这种趋同的指望放在哲学家最终会接受他所偏爱的那种认知主义和基于价值的立场。这是消除分歧或说去掉心头之刺的另一种方式:证实只有一种立场是我们可以理性地接受的。

无论是通过理论的趋同还是通过决定性证实竞争观点的不足(inadequacy),汲汲于消除分歧是帕菲特著作的鲜明特征。有时候它表现为一种紧迫感。他对沃尔芙的回应就体现了这一点。沃尔芙把帕菲特看作是竭力表明“存在着单个的真道德,浓缩为单个的最高原则,不同的传统摸索着,且都是以自己独立且不完善的方式,走向它”(32)。相对比而言,她本人说道,如果“最终表明”道德没有这样一种统一的原则,“这也不是道德悲剧”(33)。帕菲特在回应中说,“如果不存在单个的最高原则,这不会是道德悲剧,但如果没有任何单个的真道德原则,这将是一个悲剧。”他补充道,“如果我们不能解决这种分歧,这就使我们有理由怀疑是否存在任何的真道德。这可能最终表明道德什么也不是,因为也许道德只是一种幻觉”(151)。也许正是前景“黯淡”的不祥之兆,乃至如其所担忧的“更黯淡”前景——根本没有任何重要之事,在帕菲特心头挥之不去,才能解释他消除分歧的那种紧迫感。无论是否赞同他对来自深刻分歧之威胁的这种估量,人们都不可能不留下深刻印象的是:他在追求目标时所表现出的那种非凡的创造力和纯粹智识的强度。他富有挑战性的大量讨论,通过与斯坎伦、伍德、沃尔芙和赫尔曼的交流之襄助而得以阐明,由此以新颖且不同寻常的眼界重塑了常见的争论,开启了许多富有成效的新思路,从而可让哲学家做出更进一步的考察。只要人们对道德、规范性和合理性理论感兴趣,他们就无法忽略这本杰出的、发人深省且有着坚定主张的著作。

xxxii

序 言

xxxiii

既然本书有内容提要，我在此就不谈内容了。本书尽管篇幅很长，但其中包含了几本小书。第三部分没有任何重要的东西依赖于第二部分，因此你可以只看第一和第三部分。你如果主要对伦理学感兴趣，那就可以只看第6-17章。如果你的兴趣主要在于理由、合理性和元伦理学，那么你可以只看第一和第六部分。

在说明如何写作其伟大而乏味的《伦理学方法》时，西季威克谈到，他有两个“导师”：密尔和康德。我的两个导师则是西季威克与康德。

康德是自古希腊以来最伟大的道德哲学家。而我相信，西季威克的《伦理学方法》则是迄今为止写得最好的伦理学著作。有些著作的成就更伟大，如柏拉图的《理想国》和亚里士多德的《尼各马可伦理学》，但西季威克的著作中重要且为真的主张最多。西季威克尽管没有柏拉图、亚里士多德、休谟和康德等人伟大，但能够写出更好的著作。这一点并不令人惊讶，因为他生活于较为靠后的年代。生活在荷马和莎士比亚时代之后的诗人或剧作家，并没有任何相对于前者的优势；但哲学不一样，哲学会进步，因而生活在后期的哲学家有优势。

无论是西季威克还是康德，都有其弱点和不足。例如，西季威克有时候非常无趣，康德则会让人发疯。承认这些弱点，说明为什么我们不该对这些弱点失望或被其所阻碍，我希望这样可以说服一些人，让他们去读或重读西季威克的《方法》及康德的一些书。

xxxiv

就做对比而言，康德与西季威克是绝配。例如，康德在讨论自己的成就时写道：

批判哲学必定非常确信：它会不可抗拒地趋向于既满足理性的理论目的，也满足理性的道德（即实践）目的，永远无需做出观点上的改变、有所动摇或要重构为另一种形式。《批判》的体系立基于完全稳固的根基之上，且是一劳永逸确立的；它会证明：对于人类在一切未来时代的最高贵目的来说，它都是必不可少的。

西季威克写道：

这本书没解决任何问题，但可能会帮助一两个人澄清思想，一点点而已。

康德极具原创性，提出了某些崇高的主张，既让人紧张又让人激动。西季威克知道自己缺乏这些品质。他对一个朋友写道：“我喜欢自我批评，对此做了如下表述：

优点：总是考虑周到，常常是精致的：一般是明智且不偏不倚的：从正确的视角（points of view）来处理问题。

缺点：无足轻重的，条理不清的：文风呆板沉闷，论证中没有任何真正醒目或创造性的因素。

西季威克还谈到他“冗长、难懂且乏味这个要命的缺陷。”

最后的言辞过于严厉了。西季威克的书尽管篇幅很长，有些章节现在也可以被忽略，但并不冗长。西季威克很少会重复，许多重要观点的提出都是简明扼要的，并且只提一次。西季威克的书也不难懂，尽管有些主张和论证是复杂的，但几乎都写得非常清楚。

对于西季威克的枯燥，需要再说几句。怀特海极为反感西季威克的《方法》，看过它之后再也没看过任何伦理学的书。但是凯恩斯读了西季威克的回忆录和通信集后，说道，“我从来没有见过这样一本书，它如此枯燥，但又如此让人着迷。”我们值得从这本书里做些引用。西季威克在讨论英格兰的教堂时写道：

xxxv

在剑桥，我逐步形成了这样的看法：它是某种曾经活着且生长着

的东西，而之所以现在还存在，只是因为它是一座结构复杂的大厦之中的某根柱子或拱壁，人们不知它是否有价值，但目前还没有人想要撤除。而在此刻，我的感觉毋宁是：我仿佛在注视一条大鱼，它跃出水面，快乐而无碍地在大路上游动。

这里还有两段：

毫无疑问，英格兰的男人主要是在不正常的时段里坠入爱河，比如读书会、海边、番邦旅馆或圣诞节，或者任何这样的场合：不管是外部环境还是内在的情绪，只要有任何因素打破了永恒的坚冰。不幸的是，若这些偶然的融化所持续的时间不够久，由之所获的一切优势都会消失。因各种原因而相交的两条生活线也许就会永远相隔，由此是雪上加霜。

我养尊处优，却承受着人性的负荷，其结果是难负其荷。归根结底，帕斯卡讲得几乎没错：一个人如果抱有无尽的怀疑，如果这种状况如以水洗胃，如油之浸骨，那就应当身处空荡荡的地下室，披麻抹灰，悲苦不堪；而不是品着47年的葡萄酒，与克拉克作高雅之谈。走进自己的房间，我觉得很陌生，很可怖。我给你写信的原因即在于此。但在此再说一句——一个人如果允许“光阴恨短”的念头滋长而过于强烈，那么看来就会将其一辈子浓缩成狂热的片刻。

要么这个世界感到了我的脉搏，要么我就死了。

想想看，说这话的人死了，是个二流货色——对于所有像这样的人，有谁在意？

蝴蝶尚且惧死。

对我而言，这种情绪很陌生。但我今天在特兰平顿干掉了一只蜘蛛，并说，“这是感受”。除了体验感受之外，我会是什么？

西季威克也会打趣，他的谈吐被形容为：“就像一条波光粼粼的小溪，带有似乎会发光的波纹。”不过《方法》的第一稿就只有几个笑话，后来还删掉了。然而，本书的绝大部分都写得很好。

例如：

假定……“只服从个人自己”这个理想可以被代议制民主实现（即使是大致上的），这种看法的荒唐就显得更明显。因为代议院通常只是由国家的一部分人选出来，而每条法律也只是由议院里的一部分人所赞同；若某条法律得到多数议员的赞同得以通过，而某人所选议员反对，那么说“此人已经同意这样一条法律”就是荒谬的。

更冷静的：

由此义务的秩序（Cosmos）实际上归于混沌，人们为了构架理性行动的完美理想而做出的长期的智识努力，看来命中注定会失败。

这一令人印象深刻的严峻主张有着某种康德式的力度，有关康德的另一段也是如此：

我不能基于这种资源来认为：处于道德必然性之下就是把我的一切义务看得有如它们是上帝的命令，尽管我并没有资格推测确实存在任何这样的最高存在者（Supreme Being）。我完全感觉不到要迫于实践的目的去相信找不到任何根据要作为思辨真理来接受任何东西，我甚至无法想象这些语词看来要描述的那种心态，除非是作为某种暂时且愚蠢而失去理智，从而陷入对哲学的狂热之途。

xxxvii

好的段落有很多，但太长无法完全引用，有一段是这样结尾的：

自私者会错过由广泛兴趣所带来的崇高感和升华感，还会错过不断参与（指向比个人幸福更稳定的目的的）某些活动带来的更为安稳和宁静的满足：他会错过尤为丰富的美妙感受，这种感觉取决于同情的综合回报，通常可见于我们为爱人或恩人效劳时……他会以无数的方式感觉到，在自我生命的节奏与更大生命的节奏之间的不协调——前者只是后者之中微不足道的片段。

另一段的结语是：

某人说“邪恶啊，你就是我的善”，并依此行动——即便是这样一个人，对其行动之中可怕的不理性也只会有一种模糊的意识——这种模糊是由于如下的虚妄想象：此刻他已经走到人生末途，能让人钦佩的唯一机会必定在于坦率而一贯的邪恶。

西季威克提醒他的朋友们：他的书致力于“思想的精确”，因而“必定会有点枯燥和让人讨厌。”但这种精确通常表达得非常好。例如，西季威克在讨论友谊时写道：

一切强烈的亲密情感都是常识所赞赏的，同情则不是如此；常识不满于上述情感的衰退，同情的消退则没有那么令人遗憾。

xxxviii 许多句子尽管枯燥，却又有那么点反讽的味道。例如：

人们可能说孩子应该感激其亲生父母。但是仅赋予生命而不提供任何让生活幸福的必需品，看来只是一种可疑价值的福音，当这种福音的降临根本就不考虑接受者时，它极少能够激起人们的感恩之心。

……A比B更幸福若只是因为超出其控制的环境使然，那看来就没有任何不义之处。

因此谨慎地表达的效用主义结论看来是这个样子的：保密才会使一个行动正确，而不保密则不是如此，这种意见本身就应该保密起来；同样，秘传道德是有利的这一学说本身应该保持为秘传，这样才是有利的。

……真正具有穿透力的批评，特别是在伦理学之中，要求带有同情性的耐心努力——这是布莱德利先生从来都学不会的，还要求一种平和的心态——这是他从来都无法保持的。)

[这本书]看来是摧毁性的，但是由于过度争议失败了。这种破坏性攻击中至少应该有一种假模假样的公平。

西季威克其实是在进行颠覆之时，其反讽让他看似一本正经。例如，威廉

斯就受到了误导，他对西季威克对性道德的讨论这样写道，尽管有时候稍有冒险，但西季威克“对贞操概念的使用是极没有批判性的”。西季威克确实问过：“那么贞操会禁止什么样的行动呢？”但我们如果更仔细地阅读就会发现，他的答案是没有任何行动。在1874年英格兰出版的一本书中，主张为了性快乐自身的缘故而沉浸于其中在道德上完全无可指责，虽然用词谨慎，但显然不止是“稍有冒险”。

当人们发现西季威克枯燥乏味时，他们常常回应的不是其文风，而是他最伟大

xxxix

一直在读孔德和斯宾塞的东西，对他们的智识力量和勤奋早就非常钦佩，不过更多的是惊讶于他们自满的那种自信。在我看来，他们都不知道自我批评为何物。我很好奇这是不是与其成就不可分离的一种缺陷。我确定自己的自我批评阻碍我得到积极且鼓舞人心的作品：但另一方面，我感到，如果我的作品还有点价值的话，这也归功于它。

西季威克特别善于发现对其观点的反驳所具有的力量。詹姆斯听了西季威克对一篇论文的辩护后，他这样说道：

西季威克所展现的那种反思性坦率，有时候让人发怒。没有人有权对其对手如此公平。

例如，在讨论对手的一本书时，西季威克写道：

我将尽可能是赞扬它……此书的作者是很有品味的……但是，在我看来，这好象完全不是他写的书：我几乎没法以恰当尊重的态度来对待他的理论。无疑，在他看来我也是如此：我们都是对的吗？这本书使得我对伦理学极为沮丧。

这些美德使得西季威克很难读。正如布劳德所解释的，一个问题是西季威克：

不停地提炼、限定、提出反驳和回应它们，然而又发现对答案的进一步反驳。这些反驳、驳斥、答覆和再答复的每一个本身都是可敬

的，不停地有助于作者的准确和公正。但是读者容易失去耐心，会失去论证的主线，当他从桌子旁站起来时发现，他很钦佩地读了很多，但所获无几。

xi 在某种程度上讲，我们第一次读《方法》会是最糟糕的，因为书中没有什么醒目或激动人心的东西。但每次重读的时候，我们都会注意到一些好的新颖论证，而这是以前所忽略的。至少这是我的发现。

西季威克在再次批评自己时写道：

我并不是一个有原创性的人，我对自己作品的评价越来越低。

这些话也过于严厉。西季威克在几个方面是有原创性的。但这不是使他伟大的地方。康德和休谟等其他的哲学家更有原创性，也更出色。这些哲学家像牛顿和爱因斯坦一样是那种最显而易见的天才。西季威克更像达尔文，他所具有的是那种被称为“强化到近乎天才的良好感觉”。正如布劳德所论，在《方法》中，“几乎所有主要的伦理学问题都得到了极精确的讨论。”西季威克在很多事情上都是正确的。他对古代和现代伦理学中最主要的三个主题（利己主义、享乐主义和后果论）都给出了最好的批判性论说。他书中最长的四个部分，也对非后果论的多元主义常识道德给出了最好的批判论说。尽管西季威克也犯错，我在脚注中提到了一些，但我认为，他的错并不多。这些事实使得《方法》这本书成为了每个对伦理学感兴趣的人最好是要读和记的书，且能假定其他人都读过了。

我对于西季威克的感激很容易描述。我成为哲学研究生的理由之一是这个事实：在想知道要如何度过这一生时，我发现很难决定什么是真正重要的事情。我知道哲学家致力于回答这个问题，从而变得明智。不过我非常失望地发现，无论是教我的哲学家，还是被告知要去读的哲学家，他们都认为“什么重要”这个问题并没有真正的答案，有的哲学家甚至认为这个问题没有意义。我买了西季威克的一本二手书，我发现他至少相信有些事情是重要的。我正是从西季威克那里得知道德哲学家应该追问的大部分问题，还有一些问题的答案。

xli 我现在转向我的另一位导师：康德。当我首次在1960年代阅读康德的《道德形而上学奠基》一书时，我发现这本书非常吸引人但也很难懂。当我三十年后重读这本书以及康德的其他书时，我出乎意料地沉迷于康德的伦理学。我在

接下来的两三年期间基本上没想过其他的东西。

值得承认的是，尽管我沉迷于康德给了我很大的活力，但这种活力一开始几乎完全是消极的。我并不怀疑康德是个天才。但我发现，自己像许多其他人一样，既非常反对康德的主要主张，也非常反对他研究哲学的方式。我会提到是什么使得我如此反对康德，以及这种态度是如何转变的，这也许会劝说某些人不再像我以前那样，对康德近乎不闻不问。

尽管康德具有西季威克所缺乏的某些重要品质，但反之依然。西季威克的写作很清楚，具有整体上的一致性，很少犯错。这些品质都是康德所不具备的。

完全不像我们第一次阅读西季威克的《方法》时的感受，在某种程度上说，第一次阅读康德的《奠基》是最好的。书里一些主张醒目且激动人心，我们并不为看不懂而担忧。但当我们重读《奠基》时，许多人会变得沮丧，从而放弃。我们判定，康德尽管是个伟大的哲学家，但不适合我们。

第一个问题是康德的文风。正是康德使得糟糕的写作在哲学上变得可接受。我们再也不能指着其他人写的极烂的一句话说，“如果任何一个人这样写作，怎么会值得我们去读呢？”总是可以这样回答：“康德又如何呢？”

还存在着更深刻的问题。当我沉迷于康德时，我试图更清楚地重新表述康德的某些主要观点和论证，结果发现这个任务很打击人。我想以融贯的方式把康德的主张串联起来，然而，康德的许多论证看来显然是无效或不合理的。要是知道即使是康德最伟大的钦佩者也有类似的感受，那对我会有所安慰。例如，奥尼尔（Oenra O'Neill）称《奠基》是康德书中“最令人气愤的”一本。

如果我早知道康德并不具有单一且融贯的理论，这也会有所安慰。当我们询问康德是接受还是拒绝某个主张时，答案往往是：既接受又拒绝。正如史密斯（Kemp Smith）所写，“援引单个落段是很不可靠的”。例如，尽管康德写过“个人在每一个时刻的义务是尽其所能做最好的事情”，但他并不真的像这个主张所蕴含的那样是个行动后果论者。罗尔斯在尽力理解康德的文本时写道：“我假定永远不存在明显的错误，无论如何不会是重要的错误。”但是必定会有这种错误，因为康德提出了许多相冲突的主张，这些主张不可能都为真。正如史密斯所指出的，康德经常“直截了当地自相矛盾”，并且“几乎每一个专业术语的用法都有着不同且相冲突的意义。他是伟大思想家中最不精确的”。（为了避免刺激黑格尔主义者，我们也许应该加上“最不精确之一”。）

康德写过，“一致性是哲学家最大的义务”。并非如此。原创性和清晰性至少是同样重要的。康德的伟大主要在于他具有许多原创性且富有成效的观念。

如果康德总是保持一致的话，他不可能具有全部的这些观念。

当我第一次重读康德时，我发现最令人气愤的不是康德的模糊性和不一致性，而是一种特殊的言过其实、虚假的修辞。例如，康德写道：

我们如果回顾以前在发现道德原则上一直所做的一切努力，就无需惊奇为什么它们必定全都会失败。它们看到了人类要受约束于依据其义务的法则，但从未想到：个人只服从由其自身所给予的而依然普遍的法则。他有义务仅以一致于自身意愿的方式行动……

我并不介意第一个句子中的夸张。我们可以低调一点，把“一切……必定会”转变为“它们中的一些确实会”。但由于我们知道康德相信绝对命令，所以我惊讶于康德的第二个句子。我问一个康德主义者，“这意味着，如果我不把康德的命令作为自己的法则，我就不受它管辖吗？”我得到的答案是，“不，你必须为自己立法，并且只有一种法则。”这个回应让人发疯，就像旧苏维埃所宣扬的所谓“人民民主”：其中选举是强制性的，并且只有一个候选人。当我说“我并没有把康德的命令作为自己的法则”，我被告知“不对，你已经这样做了”。这个回应更为糟糕。对于这类主张的恼怒，本书可能仍有所流露。

然而，正如我说过的，这种恼怒已经消失。我现在读了康德其他的书，我知道了那种使得康德做出那些蛮横主张的激情。在较为平静时，他做出了更好的其他主张。例如，据说康德曾讲过：

自杀是激发恐惧与憎恶的最可鄙罪行……一个如此完全不尊重自己生命的人……没有办法约束他不去犯最可怕的罪恶……

但他也说过：

斯多葛学派有关自杀的原则之中不乏灵魂的高贵：我们可以放弃生命，如同逃离烟雾弥漫的房间。

不仅如此，康德有些热情洋溢的论证也极有魅力。当谴责自杀时，康德说道：

如果自由是生命的条件，那它不可能用来结束生命……生命被假

定用来导致无生命，但这是一个自相矛盾。

正是“假定”这一用词在这里显得是如此的可爱。一个评价者表示，自杀包含着一种矛盾，因为基于康德的观点，正是我们赋予自己的目的以价值。如果我们杀死自己以避免痛苦，我们就

切掉了这种目的的价值根源——它就根本不再是一个真正的目的，追求它也就不再是合理的。

这个结论来得太迟了。

对于其他的此类例子，我们可以考虑康德的这个主张：如果我们撒谎，xliv“哪怕是为了实现某个真正的好目的”，我们也就“违背了在我们自己人格中的人性尊严”并且使得自己“成为纯粹虚有其表的人”，他“具有的价值比纯粹的物件还要低”。我们应该忽略这种情感的爆发。但就在接下来的另一页，康德表示，如果一个作者问我们是否喜欢他的作品，可以允许我们去迎合他。

康德有时候被认为是一个冷酷、枯燥的理性主义者。但他其实是一个情感极端主义者。正如西季威克所写，“哦，我是多么同情康德，同情他既怀着对综合的热望，又要接受其理性批评的宣判”。康德极少使用诸如“大部分”、“许多”、“几个”或“一些”这样的语词，而是偏爱使用“全部”或“全不”来写作。他说，康德使用“善”意指“实践上的必然”。他极少使用“一种理由”（仅仅是算作支持某个行动的事实）这个概念，因为他偏爱的规范概念是被要求的、被允许的和被禁止的。从性情上讲，我也是一个极端主义者，总是竭力变得更像西季威克。

牛津大学曾经有一种有用的评分：Alpha Gamma（最好与最差）。正如每个人都会同意的，康德的书是纯粹的Alpha Gamma，没有任何beta或平庸。我们的分歧只在于康德的作品中有多少是Alpha，多少是Gamma。如果我们发现了Alpha的东西，我们就可以忽略Gamma的东西。

我认为，康德的有些观点极为接近休谟的。康德是个更危险的反智主义者，因为不像休谟，他看来提升到了他称为的“纯粹理性”。我认为，康德的影响在其他某些方面是很坏的。但他非常伟大，他的影响在其他较不明显的方面一直是好的。尽管康德提出了很多错误的主张，许多论证也是失败的，但是他给了我们深刻的真理。我像西季威克一样，有时候发现他“完全是个启示”。康德

的书极具思想激发力。正如罗尔斯所写的：“我们研究的这些作品，其与众不同的绝妙之处部分地在于，它们对我们言说所能达到的深度以及多样的言说方式。”

xlv

在本书中我试图对康德的最高道德原则的绝大部分表述有所讨论。尽管本书的论证主要在第一、第三和第六部分，而第二部分的主要部分就是为那而写的。除了极少的章节（主要在第二部分和附录）外，我并不讨论康德观点的细节。

我现在转向我受惠最多的其他人。当我年轻时，绝大部分哲学家相信不可能存在任何规范真理。绝大部分的经济学家、社会科学家和更广泛的西方世界中的人也是这样。受过良好教育的非宗教人士理所当然地接受事实（这是客观的）与单纯价值之间的区分。现在的情况几乎没有什么改变。当某个经济学家最近主张他的建议并不包含任何价值判断时，有个人说：“不对，它们包含了。你假定了我们应当去做这样的事情，它不会使任何人变坏，但会使某些人变好。”经济学家答道：“那不是价值判断，因为每个人都接受它。”

西季威克早已离世，在其时代尚能找到对实践哲学和道德哲学抱有更大希望之人，受此激励，我也要找几位抱有如此希望的在世哲学家。内格尔，特别是他对理由和不可还原的规范性真理之主张，对我的鼓舞最大。我也从斯坎伦那里学到了很多，以致都搞不清有些思想究竟是他的还是我的。我把这本书献给他们俩。

我也感激科丝嘉，正是她令人印象深刻的书让我重读康德，正是她称为“独断的理性主义”的批评使得我从自己不独断的昏睡中醒悟过来。我也很是得益于一系列卓越地讨论康德或为其所鼓舞的书和论文（即使还不够），其作者是Henry Allison, Marcia Baron, David Gauthier, Richard Dean, Jeffrey Edwards, Stephen Engstrom, Paul Guyer, Barbara Herman, Thomas Hill, Samuel Kerstein, Patricia Kitcher, Onora O'Neill, Thomas Pogge, Andrews Reath, Jerome Schneewind, David Sussman, Roger Sullivan, Allen Wood。

xlvi

我还受益于许多其他人，他们对早期手稿做了评论，由于无法在短短篇幅中道尽受益的方式，我只能向这些人表示特别的感谢。

对此书的所有部分都做了评论的那些人，我最感激下述人士：Robert Audi, Selim Berker, Talbot Brewer, John Broome, Ruth Chang（我把第16章献给他），Eugene Chislenko, Jerry Cohen, Garrett Cullity, Jonathan Dancy, David Enoch, William Fitzpatrick, Shelly Kagan, Guy Kahane, Niko Kolodny, Michael Otsuka, Ingmar Persson, Jacob Ross, Kieran Setiya, Larry Temkin。书的某些部分是与这些人合作的。

我也极得益于以下人士：Marcello Antosh, Richard Arneson, Rudiger Bittner, Mary Coleman, Roger Crisp, Stephen Darwall, Harry Gensler, Reto Givel, Elizabeth Harman, Brad Hooker, Frances Kamm, Joseph Mendola, Jefferson McMahan, Liam Murphy, Leonard Katz, Robert Myers, Martin O'Neill, Douglas Portmore, Stuart Rachels, Peter Railton, Karl Schafer, Samuel Scheffler, Michael Slote, Saul Smilansky, Jussi Suikkanen, Stephen White。

只给第一个部分做了评论的那些人中，我受惠最甚的是下列人士：Melissa Barry, David Copp, Joshua Gert, Pamela Hieronymi, Julia Markovits, Sven Nyholm, Connie Rosati, Jeffrey Sebo, David Sobel, Sigrun Svavarsdottir, David Velleman, Michael Zimmerman。

评论了我有关康德的看法的人之中，我受惠最甚的是下述人士：Marcia Baron, David Cummiskey, Richard Dean, Jeffrey Edwards, Paul Guyer, Thomas Hill, Samuel Kerstein, Patricia Kitcher, Thomas Pogge, Allen Wood。Edwards, Kitcher和Pogge就我对康德的诠释所做的评论，还有Samuel Freeman和Leif Wenar对我有关罗尔斯的主张所做的评论，我并未做出足够的回应。

只对第六部分做过评论的那些人，我受惠最甚的是下述人士：Robert Adams, Paul Boghossian, Laurence Bonjour, Nicholas Bostrom, Philip Bricker, Justin Clarke-Doane, Terence Cuneo, Cian Dorr, Kit Fine, Stephen Finlay, Alvin Goldman, Bob Hale, Michael Jubien, Thomas Kelly, Brian Leiter, William Lycan, Tim Maudlin, Brian McLaughlin, Charles Parsons, Simon Rippon, Stephen Schiffer, Mark Schroeder, Russ Shafer-Landau, Peter Singer, Knut Skarsaune, Robert Stalnaker, Brian Weatherson, Ralph Wedgwood, Timothy Williamson。

xlvi

我还受惠于下述人士：Larry Alexander, Henry Allison, Gustaf Arrhenius, Elizabeth Ashford, Bruce Aune, Annette Baier, Matthew Bedke, Akeel Bilgrami, Daniel Boisvert, Matthew Boyle, Sarah Buss, Krister Bykvist, Thomas Carson, Timothy Chappell, Daniel Cohen, Joshua Cohen, Robert Curtis, Gordon Davis, Paul Dinkin, Thomas Donaldson, Dale Dorsey, Jamie Dreier, Julia Driver, Jerry Dworkin, Andrew Egan, Nir Eyal, Geoffrey Ferrari, Claire Finkelstein, Katrin Flikschuh, Johann Frick, Jerry Gaus, Berys

Gaut, Tamar Gendler, Pablo Gilabert, Margaret Gilbert, George Giovanni, Joshua Glasgow, James Grant, Liron Greenstein, Alex Gregory, Ish Haji, Jason Hanna, Robert Hanna, Joshua Harlan, Daniel Hausman, Allan Hazen, Christopher Heathwood, Dieter Henrich, David Heyd, Alison Hills, Nathan Holcomb, Mike Huemer, Thomas Hurka, Paul Hurley, Susan Hurley, Frank Jackson, Dale Jamieson, Justin Jeffrey, Leonard Kahn, Robert Kane, Stephen Kearns, Paul Klumpe, Richard Kraut, Rahul Kumar, Joel Kupperman, Arto Laitinen, Robin Lawlor, Mark LeBar, James Lenman, John Leslie, Hallvard Lillehammer, Don Loeb, David Lyons, Tienmu Ma, Jacqueline Marina, David McCarthy, Kris McDaniel, Dennis McKerlie, Chris McMahan, David McNaughton, Elijah Millgram, Adrian Moore, Sophia Moreau, Adam Morton, Istvan Musza, Jan Narveson, Stephen Nathanson, William Nelson, Michael Neumann, Kenneth O'Day, Avner Offer, Onora O'Neill, Serena Olsaretti, Jonas Olson, Toby Ord, Leah Orent, Francesco Orsi, David Owens, Stephen Palmquist, Herlinde Pauer-Studer, David Phillips, Christian Piller, Richard Price, Bogdan Rabanca, Wlodek Rabinowicz, Toni Rnow-Rasmussen, Joseph Raz, Andrews Reath, Bernard Reginster, Michael Ridge, Arthur Ripstein, Michael Rohlf, Gideon Rosen, Mike Rosen, Carol Rovane, Angelica Rudenstine, Julian Savulescu, Jerome Schneewind, Dieter Schoenecker, Frederick Schueler, Bart Schultz, Sally Sedgwick, Jeffrey Seidman, Matthew Seligman, Julius Sensat, Andrew Sepielli, Robert Shaver, Walter Sinnott-Armstrong, John Skorupski, Holly Smith, Michael Smith, Tom Sorell, Carlos Soto, Amia Srinivasan, Cynthia Stark, Philip Stratton-Lake, Galen Strawson, Bart Streumer, Roger Sullivan, Adam Swenson, Folke Tersman, Jens Timmerman, Torbjörn Tännsjö, Pekka Vayrynen, Edna Ullmann-Margalit, David Velleman, Benjamin Vilhauer, Gerard Vong, Alex Voorhoeve, R. Jay Wallace, James Walmsley, Paul Weirich, Kenneth Westphal, Evan Williams, Chris Woodard, Helena Wright, Masahiro Yamada。

xlviii

我也要感谢在写作本书的这么多年间，万灵学院的讲座教席给我提供的诸多优待。我还感谢唐纳基金支持我把讲座扩展成第二和第三部分。我既要感谢写了本书第四部分的各位评论者，还要感谢弗勒的编辑工作。最后我要感谢牛津大学出版社的Peter Momtchiloff，这么多年来他给了我很多合理的建议。

第一部分 理由

第一章 规范性概念

1 规范性理由

我们是既能理解又能回应理由的动物。当事实算作支持我们具有某种信念或欲望，或支持我们以某种方式行动时，它们就给了我们理由。当我们做某事的理由比做其他任何事情的理由更强时，这就是我们最有理由（most reason）做的，是我们应该（should）、应当（ought to）或必须（must）做的。尽管是事实给了我们理由，但我们能理性地想要或去做的事情反过来取决于我们的信念。

2 涉及理由的善性（goodness）^①

事物可能具有某些可好可坏（good or bad）的特征，这些特征给予我们理由以某些方式回应这些事物。事物可能对于某些特定的人们是好的或坏的，也可能在蕴含理由的意义上是非个人地（impersonally）好的或坏的。基于某些被广泛接受的理由观，任何事情都不可能以这些方式是好的或坏的。

① Good（goodness）的翻译没有强求一致，而是根据表达汉语的习惯，当它泛指一般时译为好的，道德味道浓时则译为善（善性）或价值，物质味道浓时则译为物品、益品、利益等。

第二章 客观理论

3 两种理论

按照主观理论，我们最有理由做的就是能够实现或满足我们当前欲望或目标的任何行动。有些主观主义者诉诸我们实际的当下欲望或目标；还有些人诉诸这样的欲望或目标，即如果我们仔细地考虑相关的事实，我们将会具有的欲望、目标或会做出的选择。既然这些事实都是有关我们的，我们就可称这样的理由为主体给予的（subject-given）。按照客观理论，仅当或因为我们所做的事情或试图获得的事物在某种程度上是好的或值得做的，我们才有理由那样去做。既然这些都是有关这些欲望或目标之对象的事实，我们就可称这样的理由为对象给予的（object-given）。它们也是基于价值（value-based）的。这两类理论通常具有深刻的分歧。我将论证，我们应该接受某种基于价值的客观理论。

4 回应理由

当我们认识到事实给予我们很强的理由让我们具有特定的欲望时，我们对这些理由的回应很少是自愿的。我们也无法选择如何回应我们要具有特定信念的大部分理由。我们的理性在部分程度上就在于非自愿地回应这些理由。

5 状态给予（state-given）的理由

我们如果拥有某些信念或欲望就会是好的，这似乎可以给予我们拥有该信念或欲望的理由。但这种理由根本无足轻重。

6 享乐式^①理由

相同的事实给予我们对象给予的理由既要具有某些欲望，也要努力满足它们。我们想要的总是某种有可能的事件，它在广义上涵盖行动和事态。我们具有目的式（telic）理由想把某些事件作为目标，或为了事件自身之故而想

① Hedonic译为享乐式的，指与快乐和痛苦等有关的，是描述性的，hedonistic译为享乐主义的，意指重视快乐与痛苦这种感知的，是评价性且是具有正面意义的，hedonism则译为享乐主义。要注意的是，国内一般对享乐主义持有贬义，但帕菲特这里显然没有这种消极的意义，反而具有极正面的意义，阅读时请注意。

要这些事件，也有工具性理由想要把某些事件作为获得某些好目的的手段。我们最有理由做的是能够实现我们最有理由想要之目标的任何事情，因为这些目标的内在特征使其在相关意义上是最好的。

我们疼痛时，坏的东西不是我们的感知，而是“我们具有我们所厌恶的感知”的这种意识状态。类似地，我们具有我们喜好的感知是好的。这种享乐式喜好（likings）与厌恶（dislikings）不可能是合理的或不合理的，因为我们没有任何理由来喜欢或不喜欢这些感知。我们还可以对自己或他人的快乐与疼痛具有元享乐式（meta-hedonic）欲望。这种欲望或偏好可以是合理的或不合理的，因为我们可以有很强的理由要持有它们。正是我们的享乐式喜好与厌恶，而不是我们的元享乐式欲望，才使得这些意识状态是好的或坏的。因此，快乐与疼痛的例子并不支持这种观点：我们的欲望给予我们理由，并且能使得其对象是好的。

7 不合理的偏好

如果我们想把某个事件作为目标，但这个事件的内在特征给予我们很强的决定性理由想这个事件不发生，那么我们想要这个事件就是与理性相冲突的，从而是不理性的。例如，宁愿明天有一个小时的剧痛，而不愿今天有一分钟的微痛，这就是不理性的。这些主张看来太过明显而不值一提。但有些伟大的哲学家却否定这样的主张，那些接受理由的主观理论的人则不能做出这样的主张。

第三章 主观理论

8 理由的主观主义

主观主义有几种形式。主观主义可以诉诸我们当下全部的目的式欲望，或者只诉诸基于真信念的欲望，或者诉诸完全知情的^①欲望。有些主观主义者则诉诸我们经过知情且理性的慎思之后所做的选择。有些客观主义者则诉诸我们在这种慎思之后若理性就会做的选择。这些主张尽管看起来是类似的，但实际上完全不同。主观主义者仅仅主张我们应该只以程序上（procedurally）合理的方式进行思考和选择。客观主义者所做的主张是如果我们是实质上（substantive）理性的话，我们会做出的选择。按照客观主义者的观点，我们

① Informed desire在中文里很难找到一个合适的翻译，其意思是指我们在具有相关信息条件下形成的欲望，这里不得已翻译为“知情欲望”，并且一般把fully informed译为充分知情的，意指完全了解相关情况的，或具有充分信息的情况下，有时候会译为有充分信息的等，阅读时请注意。

应当理性地选择的东西取决于我们的理由。按照主观主义者的观点，我们的理由取决于我们在这种慎思后事实上会选择的东西。

9 为什么人们接受主观理论

4 既然有这么多人相信所有实践理由都是基于欲望、目标或选择的，真实情况如何能如客观理论所声称的那样，没有任何这样的理由存在呢？这些人怎么会都犯这样的错误呢？由于我们的欲望或目标看起来有几种方式给予我们理由，故也有几种可能的解释。

10 分析性主观主义

有些主张看来是实质性的，但不过是隐藏的同义反复，不管人们相信什么都可以接受它们。有几种主观主义者在主观主义的意义上使用“理由”、“应该”和“应当”，不过这些人的理论并没有做出实质性的主张。

11 痛苦论证

实质性的主观理论会具有讲不通的含义。例如，这些理论蕴含着，我们常常没有任何理由想要避免将来某个时期的剧痛。有些主观主义者会诉诸有关程序合理性的主张来回应这种反驳，但这种回应是失败的。

第四章 进一步的论证

12 全部或全不论证

主观理论还可能蕴含着：我们有决定性的理由为了痛苦自身的缘故使自己处于痛苦之中，浪费我们的生命，以及获得其他坏的或无价值的目标。为了回应这种反驳，主观主义者也许会主张，因为某些欲望或目标会给我们一种理由，所以我们必定有某种理由持有这种欲望或目标。但这些人提出的这种主张是得不到捍卫的。基于主观理论，重要的不过是某个行动是否会实现我们当前完全知情的欲望和目标。我们想要什么或努力想获得什么是不相关的。要么这些欲望全部都给我们行动的理由，要么没有任何欲望能给我们理由。既然有些欲望显然不能给我们理由，我们就应得出：没有任何欲望能够给我们理由。

5 人们可以主张，有些欲望可以给我们理由想要其他的欲望，但任何基于欲望的理由链条必定始于我们没有任何理由想要的欲望。既然这样的欲望不可能有证成地主张给我们理由，主观主义者不可能有证成地主张我们有基于欲望的理由具有任何欲望和目标，或者以任何方式行动。

13 不融贯论证

许多主观主义者主张，我们最有理由实现的不是我们实际的当下欲望和目标，而是在我们知道相关事实条件下我们会具有的那种欲望和目标。这些人还主张，当我们做出重大决策时，我们应该尽力对我们行动不同的可能结果有更多的了解，这样我们就将具有更好的知情欲望。既然主观主义者否认这些结果的内在特征会给予我们理由，他们无法融贯地提出这种主张。

14 理由、动机和福利

我们若是主观主义者就必须否定事件可以对特定的人是好的或坏的，或者在蕴含理由的意义上是非个人性地好的或坏的。当某些作者主张有些生活对某人是最好的时候，他们意指，这种生活是这个人充分知情时做了程序上合理的慎思后事实上会选择的。基于某种论说，对某个人最好的生活可以是一种完全充满苦难的生活。这不是一个有用的主张。其他的论说则以其他的方式而失败。

15 支持主观主义的论证

基于主观理论，没有任何事情是重要的。对于支持这种黯淡观点的论证，我们应予以拒绝。

第五章 合理性

16 实践合理性与认知合理性

如果我们能够很好地对理由或显见（apparent）理由做出回应，那么我们就是理性的。如果我们具有相关事实的信念，且这些事实为真就可以给我们某种理由，那么我们就具有了显见理由。如果我们的信念若为真则我们就有充分的理由持有某些欲望和进行某些行动，那么这些欲望和行动就是合理的。有些人还会加上，我们的欲望和行动要是合理的，它们还必须依赖于合理的信念。这个主张是误导性的，不值得做这样的主张。

基于某种观点，认知合理性的独特性在于获得真信念这个目标。还有一种更好的观点。除了在认知合理性和实践合理性之间做出更深刻的区分外，我们应该以不同的方式且在不同的地方做出这种区分。

17 有关理由的信念

按照有些作者的看法，要想成为完全理性的，我们并不需要对理由或显见

理由做出回应。只要避免某种不一致——比如不要回应我们自己相信是理由的因素——就足够了。这种观点太狭隘了。

18 有关合理性的其他观点

我们欲望的合理性并不如许多人所主张的那样，取决于这些欲望是否一致、我们如何获得这些它们或者我们具有这些欲望是否有好的效果。如果我们的欲望所依赖的信念若为真就会使得欲望的对象或我们想要的东西在某种意义上是好的或值得获得的，那么这样的欲望就是合理的。

第六章 道德

19 西季威克的二元论

西季威克似乎主张，我们能从两种视角（points of view）来评估我们理由的力度。我们若从个人视角来看，则自利理由是最高的；而我们若从不偏不倚的（impartial）的视角看，则不偏不倚的理由是最高的。为了比较这两种理由，我们就需要第三种中立的视角。既然没有这种视角，自利理由和不偏不倚的理由就是完全不可比的。这两类理由若发生冲突，没有哪一方能是更强的。不管是做不偏不倚地最好的事情，还是做对我们自己最好的事情，我们总是有充分的或不可挫败的理由。

7 我们应该拒绝西季威克的论证。我们应当从我们实际的个人视角来评价一切理由的强度，我们并不需要一种中立的视角。我们也应该修正西季威克的结论。我们持有偏向性的个人理由要特别关注的不仅是自己的福祉（well-being），还有其他某些人（比如我们的近亲还有我们所爱的人）的福祉。我们与这些人具有紧密的纽带。我们也有不偏不倚的理由关爱任何人的福祉，不管这些人与我们的关系如何。尽管存在着关于这两种理由的相对强度之事实，但西季威克的观点在部分程度上是正确的，因为这种比较即使在原则上也是极不精确的。正如广义的基于价值的客观理论所主张的，当两个可能的行动之一是不偏不倚地最好的，另一个则是对我们或对与我们具有紧密纽带的人更好，我们通常具有充分的理由来做两者之一。

20 最深刻的问题

我们除了能问“我们最有理由做什么”之外，还能问“我们在道德上应当做什么”。如果这些问题通常具有冲突的答案，那么我们就最有理由做不当的行动，道德因此就会被破坏。像其他的规范要求一样，只有当道德要求给我们理

由时，它们才是重要的。

尽管理由是最根本的，但接下来的许多内容是有关道德的。但是我也将讨论理由。几种道德原则和理论诉诸这样的主张，即在实际的或虚拟的情境下，我们最有理由或充分的理由同意、赞同、想要、选择和做什么。

第七章 道德概念

21 无知或错误信念下的行动

通过区分“道德上应当”和“不当的”的几种意义，我们可以认识到一些重要的道理并且避免一些不必要的分歧。行动可以在事实相对、证据相对、信念相对和道德信念相对的意义上是不当的。有关这些不当性的事实为不同的问题提供了答案。当我们应该做什么取决于我们行动的效果时，我们应该尽力去做的事情不是事实上最好的，而是在期望上最好的。

8

22 其他类型的不当性

还有几种其他意义的“不当的”，它可以指向不同类型的不当性。这些意义的大部分都是值得使用的。

正如我认为是的，是否存在某种不可还原的规范真理（其中一些是道德真理）是很难回答的。当我们在实践理由、认知理由和道德上的思考取得进展后，回答这些问题会容易一些。我不是想提出一种新的道德理论，而是想发展现存的三种理论：康德主义、契约论和后果论。

第二部分 各种原则

第八章 可能的同意

23 强制与欺骗

康德主张，如果我们对待人们的方式是他们没有可能（can't possibly）表示同意的，我们就行为不当。这个主张看似蕴含着，我们永远也不能强制和欺骗别人，因为这些行动的本质就使得人们的同意是不可能的。但这在相关意义上并不为真。

24 同意原则

康德有关同意的主张可以做两种诠释。基于给予选择原则，如果对待人们的方式是他们不可能实际上同意或拒绝同意的，这就是不当的，因为我们并没给人们能力来选择我们如何对待他们。这个原则明显是错误的。基于同意原则，如果我们给予人们能力来选择我们如何对待他们，且对待人们的方式是他们不可能理性地同意的，那么这就是不当的。这个原则更有可能是康德的意思，也有可能为真。

- 9 作为理性存在者，我们应该彼此如何相关，康德的主张对此给了一个激动人心的理想。我们也许有能力只以人们能够理性地同意的方式去对待他们，这有可能是人们应该总是如此行动的方式。

25 给予同意的理由

我们能否实现康德的理想，取决于哪些行动是人们由于拥有充分的理由同意而能够理性地给予明智的同意。如果最好的理由理论是主观理论或理性利己主义，同意原则就会失败，因为会有无数被允许甚至是道德上要求的行动是人们不可能理性地同意的。但是如果最好的理论是如我所相信的那种广义的基于价值的客观理论，那么同意原则就有可能成功。正如一些例子显示的，总是至少有一个可能的行动是每个人都能够理性地同意的。我们有理由相信，在所有这些情形中，对待人们的方式若不可能得到任何人的理性同意就是不当的。

26 一个多余的原则？

按照某些作者的看法，同意原则即使为真也没为我们的道德思想添上任何内容。道德上重要的不是“人们不可能理性地同意某些行动”这个事实，而是“给这些人以决定性的理由不同意这些行动”的各种事实。这个反驳若应用于只影响一个人的行动，它还有点力度。

但当我们的行动影响很多人时，如果只有一个可能的行动是所有人能够理性地同意的，这个事实就给我们很强的理由要以这种方式行动，也有助于解释为什么其他可能的行动是不当的，由此我们是否能够实现康德的理想就是值得追问的。

27 实际同意

如果对待人们的某种方式是人们既没有也不会实际上同意的，那么它就是

不当的。这些行动即使是人们能够理性地同意的，也依然是不当的。这根本不是对同意原则的反驳，同意原则的主张只是描述能使行动不当的事实之一。

按照某种观点，以人们实际上拒绝同意的任何方式对待他们都是不当的。这显然是错误的。看起来没有任何人能理性地同意他实际上拒绝同意的对待方式。如果这点为真，同意原则也将是显然错误的。但这个反驳是可以回答的。

按照权利原则，在未征得其实际同意的情况下，每个人都有权不受到某种方式的对待。在陈述和应用这个原则时，我们需要回答一些困难的问题。

28 道义式 (deontic) 信念^①

为解释同意原则并没有错误地要求某种不当行动，我们必须诉诸这个事实：这些行动之不当在于他处，或有其他理由。基于某种可行的假设，同意原则永远也不会要求我们行为不当，因为任何行动的不当性都会给所有人充分的理由同意我们不这样做。

29 极端要求

同意原则可能要求我们承担很大的负担，若这种要求会使他人从更大的负担中解脱出来。如果这个要求过于苛刻，我们就不得不修正这个原则。但我们依然可能会实现康德的理想。

第九章 纯粹作为手段

30 纯粹手段原则

康德主张，把任何理性存在者纯粹作为手段来对待都是不当的。若我们既把人们作为工具来使用，也如此看待他们，也就是我们会以任何能够最好地实现自己目的之方式来对待他们时，我们就是把他们纯粹作为手段来对待。基于更好版本的康德原则，把人们纯粹作为手段来对待，或者接近于这样做，这是不当的。

只要（1）我们对待人们的方式在充分重要的意义上受到了某些相关的道德信念或关怀的管辖，或（2）我们在相关意义上确实选择或会选择为这个人

^① deontic有“义务（论）的”、“道义（论）的”等译法。就本书内容看，作者是在日常道德直觉的意义上讲的。从中文语境看，“义务（论）的”或“道义论的”有较强的道德理论方面的蕴含，而“道义式的”较适合指涉常识道德直觉方面的含义。故本文一般译为“道义式”或“道义式的”。

的缘故而承担某种重大的负担，那么我们就没有把他纯粹作为手段，也没有接近于那样做。

假定某个利己主义者为了有益于自己而信守诺言，以得到所需要的帮助，或者为了获得某种回报而挽救溺水儿童。既然这个人把那些人纯粹当作手段来对待，康德的原则就会错误地谴责这些行为。我们可以限定这个原则，仅当我们的行动很可能会伤害某个人时，它才会谴责把这个人纯粹当作手段来对待。

接下来我们假定，有个无人驾驶的出轨列车正冲向一个隧道，会轧死五个人。除非你不经我同意使得我倒在地上（这会杀死我，也会阻止火车），否则五个人必死无疑。情况看来是，如果你这样做就是把我纯粹作为手段来对待。但在这个情形的某些版本中，这并不为真。我可以理性地同意受到这种方式的对待。尽管这个行动可能是不当的，但其不当性既不蕴含在纯粹手段原则中，也不蕴含在同意原则之中。

31 作为手段和纯粹作为手段

人们广泛认为，如果我们未经他人同意，把伤害他人作为实现某个目标的手段，那么我们就是把这些入纯粹作为手段来对待，这种方式使得我们的行为不当。这个观点包含三个错误。第一，当我们把伤害人们作为手段时，我们可能没有把这些入作为手段来对待。第二，即使我们把这些人作为手段来对待，我们可能并没有把他们纯粹作为手段来对待。第三，即使我们把他们纯粹作为手段来对待，我们也可能没有行为不当。

对于把人们纯粹作为手段来对待包含什么，有些人给出了其他的论说。这些论说看来要么是错误的，要么是没有用的。

32 作为手段的伤害

12 如果我们把伤害某人作为实现某个目的的手段是不当的，那么即使我们没把这些入纯粹作为手段来对待，这些行动依然是不当的。如果我们把伤害某人作为实现某个目的的手段不是不当的，那么即使我们是把这些入纯粹当作手段，这些行动也不是不当的。尽管把任何人纯粹视作手段是不当的，但我们行动的不当性从不或极少取决于我们是否把人们纯粹当作手段来对待。

第十章 尊重与价值

33 对人的尊重

我们应该尊重每一个人，但这并不能告诉我们应该如何行动。有些作者主

张，如果对待人们的方式与尊重他们是不相容的，那么这就是不当的。在困难的情形下，这无助于我们判断行动是否不当。

34 两类价值

有些事物具有一种有待促进的价值。若有关行动或事件的事实给我们理由来实现它们，那么它们在这种意义上是好的。人拥有某种应予尊重的价值。这种价值并非某种善性。

35 康德式尊严

康德使用“尊严”来指最高的价值（value or worth）。人们有时候主张，按照康德的观点，这种最高的价值只为理性存在者或人所拥有，这是一种要被尊重而不是促进的价值。但这不是康德的观点。康德主张，有好几种目的或结果具有最高的价值，并且是每个人应该去促进的目的。

康德的有些评论表明，非道德的理性具有最高的价值。但是康德的主要主张并不使他承诺这种不可行的观点。康德没有区分作为无上的善与拥有某种道德身份，而极恶的人也能拥有这种道德身份。但我们可以为康德的观点补上这个区分。

13

36 正当与善

康德主张，有些古希腊人错误地试图从有关最伟大善的信念中派生出道德法则。但康德描述了一个理想的世界，他称之为最高或最伟大的善，且主张每个人应该总是致力于产生这样的世界。康德在此似乎犯了他称之为古希腊人的“根本错误”，但情况不是这样的。

37 促进善

在康德的理想世界里，人皆有德且每个人拥有与其德性相称的所有幸福。康德主张，我们可以通过严格地遵守他的其他原则而实现这样的世界。人们常常认为，因为康德主张撒谎总是不当的，所以他是拒绝行动后果论的。情况并不是这样的。但当康德、休谟和其他人做出这样的主张时，他们并没有做出这种必要的区分。

第十一章 自由意志与应得

38 道德所要求的自由

康德论证，如果我们的行动纯粹是时间中的事件，那么这些行动就是被困

果地决定的，由此我们永远也不可能有不同的行动，道德就将是幻觉。既然道德不是一种幻觉，我们的行动就不纯粹是时间中的事件。这个论证是失败的。尽管我们只有能做其他行动才应当做不同的行动，但相关意义的“能够”是与决定论相容的。

39 为什么我们不可能应得痛苦

14 按照康德的另一个论证，如果我们的行动纯粹是时间中的事件，我们就永远不可能以那种能使我们应得痛苦的方式来对这些事情负责。既然我们能以关涉应得的方式对我们的行动负责，我们的行动就不仅仅是这种事件。尽管这个论证是有效的，但它并不健全。我们应该接受康德的这个主张，即如果我们的行动是这种纯粹事件，我们就不应得痛苦。但既然我们应该拒绝这个论证的结论，那就应该拒绝康德的另一个前提。我们的行动纯粹是时间中的事件，所以我们不应得痛苦。

第三部分 理论

第十二章 普遍法则

40 不可能性公式

康德所讲的我们的准则，其大概含义是指我们的策略及其隐含的目标。根据我们所谓不可能性公式的康德的成文版，按照不能成为普遍法则的任何准则来行动是不当的。在这一版本可能被主张为正确的含义之中，没有任何一种是有用的。

根据康德不可能性公式的真正版本，按照任何这样的准则来行动是不当的：如果每个人都接受并奉行该准则，或者每个人都相信按照它来行动是可允许的，那将使得任何人按照该准则来行动都不可能取得成功。这个公式惊人地失败，因为它并不谴责自利的杀人、伤害、强制、撒谎和偷窃。康德的公式正确地谴责许假诺。但该公式是出于坏理由去谴责这种行动，而且错误地谴责某些好的或为道德所要求的行动。

41 自然法公式与道德信念公式

康德提出了另一种好一些的公式。为了运用该公式，我们假定，我们有能力去意愿或选择“某些事情是正确的”。康德宣称，按照某些我们不能合理地意

愿其为普遍法则的准则来行动是不当的。这一普遍法则公式有三个版本。根据

自然法公式：按照某准则行动是不当的，除非我们能够合理地意愿如下情况为真：每个人都接受并尽可能地奉行该准则。

根据

15

可允许性公式：按照某准则行动是不当的，除非我们能够合理地意愿如下情况为真：在道德上允许每个人都奉行该准则。

根据

道德信念公式：按照某准则行动是不当的，除非我们能够合理地意愿如下情况为真：每个人都相信奉行该准则在道德上是可允许的。

考虑康德的自然法公式和道德信念公式就够了。这两个公式所提出的观念可表达于两个常见的问题：“如果每个人都那么做会怎样？”以及“如果每个人都像你那么想会怎样？”

在运用这些公式时，我们必须诉诸某种关于合理性与理由的观点。我们既然要问康德的公式能够取得怎样的成果，那就应该诉诸我们所认为的最佳观点。但是，我们不应该诉诸我们关于哪些行动为不当的信念，或者是这样的不当性可能提供的道义式理由，因为那样的话康德的公式将一无所成。

42 行动者的准则

康德的公式认为，某一行动的对错，取决于行动者的准则。康德所讨论的大多数准则都是或都包含策略。设想某个利己主义者仅持有一种准则或策略：“做最有利于自己的任何事。”这个人不能合理地意愿如下情况为真：每个人都奉行该准则，或每个人都相信这样的行动是可允许的。大多数利己主义者不能合理地选择生活在一个由利己主义者组成的世界，因为那会比人们奉行多种道德准则的世界糟糕得多。只要我们想象中的利己主义者奉行其准则，康德的公式就蕴含着，这个人的行动是不当的。甚至在这个人出于自利的理由去还债、加衣保暖、指望得到某种回报去救落水儿童等情况下，他的行动也都是不当的。这些蕴含显然是错误的。该利己主义者在这么做的时候，其行动没有康德

16

所称的道德价值，但并非不当。

接着考察康德的准则“从不撒谎”。康德不能合理地意愿如下情况为真：从来无人撒谎，甚至是对一个问其谋害对象藏身之处的企图谋杀者也不撒谎。因而康德的公式就蕴含着：康德如果奉行该准则而对任何人都讲真话，那他就在不当地行动。这显然是错误的。种种情况表明：在准则能够指涉策略的意义上，某一行动的正当与否，不可能取决于行动者的准则。存在许多这样的策略：奉行它们有时但并非总是不当的。行动的道德价值也不取决于行动者的准则。

康德诉诸行动者的准则，引发了其他难题。这些难题导致有些人认为，康德的普遍法则公式无法帮助我们确定哪些行动是不当的。这些人声称：在用作这样的标准时，康德的公式是不可接受的、无价值的，且无法变得行之有效。

可以使康德的公式变得行之有效。我要论证，该公式一旦以某些方式予以修正就会取得惊人的成功。

有些作者提议，康德的公式不应诉诸行动者的实际准则，而应诉诸行动者原本可以奉行的可能准则。该提议无效。

在修正康德公式的两个版本的过程中，我们应该放弃准则的概念，代之以对在考察的行动予以道德相关描述。自然法公式可变成：

我们在不当地行动，除非我们在做之事是，我们能够合理地意愿每个人在类似情形下都会尽可能做的。

道德信念公式可以变成：

我们在不当地行动，除非我们能够合理地意愿“每个人都相信这样的行动是道德上可允许的”为真。

这两个公式尚需某些进一步的修正。

17 或许有人提出反驳，如果以放弃准则概念的方式来修正康德的公式，我们就不再是在讨论康德的观点。的确如此，但这根本不构成反驳，我们是以可能取得进步的方式来发展某种康德主义的道德理论。

第十三章 如果每个人都那么做会怎样？

43 群己困境

继续讨论康德的公式会更简单，必要时再回到我们的修正版。

按照康德的 自然法公式 ，按照某一准则行动是不当的，除非我们能够合理地意愿如下情况为真：与其无人奉行，不如每个人都奉行该准则。我们通常是情况如此的群体之中的成员：如果我们每一个人而不是无人做对自己较有利之事，那么我们一起就在做对我们所有人都较不利之事。在许多这样的情况下，我们每一个人可以要么利己，要么给予他人更大的利益。我们拥有某些为道德所允许或要求的其他目标，比如说促进我们孩子的福祉，此时我们可能面临类似的群己困境。情况可能是这样的：如果我们每一个人而不是无人做较有利于自己孩子的事情，那我们就在做较不利于每个人的孩子的事情。我们不能合理地意愿如下情况为真：与其无人不如每个人都这么做。因此，如果每个人都遵循康德的 自然法公式 ，那就无人会这么做，并且这将有利于每个人。正是这样的情况下，我们拷问“如果每个人都那么做会怎样？”，可以说是最恰如其分的。

任一单个的行动所产生的副作用为极多的人所分摊，乃至对每一个人的影响都微不足道或难以察觉，对于这种情况，康德的公式尤显其价值。其中一例是我们自私地让大气过暖的行动。通过要求我们做仅为我们能够合理地意愿每个人都会做的事，康德的公式有助于我们判断在做之事会产生多大的危害，并且强有力地支持这样的观点，即这些行动是不当的。在其中某些情况下，我们可以补充说，常识道德是直接地集体自我挫败的，因而应该予以修正。

44 门槛反驳

按照某一准则行动是否不当，有时取决于有多少人这么做。存在这样一些准则：它们若为每个人都奉行就会很坏，但若为一些人奉行则是可允许或好的。“只消费而不生产食物”与“不生养孩子以便毕生致力于哲学”的准则，即为其中两例。我们大多数人不能合理地意愿“每个人都奉行这样的准则”为真。因此，康德的 自然法公式 谴责这样的行动，即便是在它们并非不当时也谴责。大多数人的准则把“他人在做什么”暗中纳入了考虑，这一事实部分地答复了这一反驳。要做出充分的答复，我们就必须修正康德的公式。

18

45 理想世界反驳

经常有人声称，康德的 自然法公式 要求我们应当仿如生活在某种理想世界那样行动，即便可以预期：在现实世界，这么做会产生灾难性的结果且明显是不当的。例如：要求我们即便在自卫的情形下也绝不使用暴力；我们被要求的各种行动方式会错误地忽视他人实际的所作所为。这种理想世界反驳可以得到答复。康德的公式没有要求这样的行动。

存在另一种问题。一旦少数人没做到我们能够合理地意愿每个人都会做的

事，康德的公式就允许余下的人做想要做的任何事。类似的反驳适用于某些规则后果论的道德理论。要答复这一新的理想世界反驳，我们应该以另一种方式修正康德的公式。这样的公式可以主张，按照某一准则行动是不当的，除非我们能够合理地意愿如下情况为真：不仅每个人而不是无人奉行该准则，而且任何其他人数的人而不是无人奉行该准则。规则后果论可以做出类似的主张。

在康德普遍法则公式的两个版本之中，道德信念公式较优。当人们反问“如果每个人都那么做会怎样？”的时候，“大多数人不会”的回答通常就够了。但当人们反问“如果每个人都像你那么想会怎样？”的时候，仅回答“大多数人不会”就是不够的。

第十四章 不偏不倚

46 黄金规则

19 康德对黄金规则的轻视是没有正当理由的。

47 稀罕性反驳与大赌注反驳

在不当地行动时，人们可能是在做某种不可能常做的事，或者是给予自己以不同寻常的巨大利益。在其中某些情况下，这些人能够合理地意愿如下为真：每个人都像他们这么做，且相信这样的行动是可允许的。因此，康德的公式就错误地允许这些人的不当行动。

48 不可逆反驳

许多不当的行动让行动者受益，但给他人施加了大得多的负担。黄金规则谴责这样的行动，因为我们不会意愿他人这样对待我们。但在运用康德的公式时，我们所问的，不是我们能否合理地意愿“其他人对我们做这样的事”为真，而是我们能否合理地意愿“每个人都对他人做这样的事”为真。而我们可能认为，即使每个人都对他人这么做，也无人会对我们这么做。在此情况下，许多犯错者能够合理地意愿如下两者均为真：每个人都像他们这么做，以及每个人都相信这样的行动在道德上是可允许的。因此，康德的公式就错误地允许这些人的行动。

这一反驳适用于许多实例。略举几例。许多男人将女人看作低人一等、拒绝承认女性的某些权利与待遇、给予女性的福祉以较小的权重，从而有利于己。要论证康德的公式会谴责这些男人的行动，我们就必须主张，这些男人不

能合理地意愿如下的任一种情况为真：他们及其他男人继续这么做以有利于己，或者是包括女人在内的每个人都相信这样的行动是正当的。由于不能诉诸这些行动为不当的信念，我们就无法为这一主张做出言之有理的辩护。因此，康德的公式就错误地允许这样的行动。类似的观点适用于如下的某些行动：许多强者或富人利用和压迫另一些弱者或穷人。

49 一个康德主义的解决方案

为免于其他种种反驳，我们应该再次修正康德的普遍法则公式。修正康德的道德信念公式就够了，它可以变成：

按照某准则行动是不当的，除非每个人都能够合理地意愿如下情况为真：每个人都相信这样的行动是道德上可允许的。

一旦每个人都相信某一行动是可允许的，那每个人就都接受允许如此行动的原则。如果某一道德理论诉诸每个人都能够合理地选择其被普遍接受的原则，那么这种理论就是契约论。因此，我们可以重述上述公式，并予之以另一个名称。根据

康德式契约论公式：每个人都应当遵循这样的原则，其被普遍接受是每个人都能够合理地意愿的。

这个公式或许就是康德一直在努力探寻的东西：道德的最高原则。

第十五章 契约论

50 合理的协议公式

许多契约论者都要求我们去想象，我们与他人拟就每个人都愿意接受的道德原则在努力达成协议。根据

合理的协议公式：每个人都应当遵循这样的原则，其被普遍接受是每个人都会合理地同意的。

契约论的这一版本要么缺乏明确的蕴含，要么给予那些拥有更强议价能力的人以不公平的有利条件。

51 罗尔斯式契约论

21 罗尔斯主张，要避免上述反驳，我们应该添加无知之幕。根据

罗尔斯的公式：每个人都应当遵循这样的原则：如果每个人都必须在不知道有关自己及其处境的任何特定事实的情况下达成协议，那么他们就都会同意，该原则被普遍接受在自利的意义上是合理的。

罗尔斯宣称，契约论的这一版本对各种形式的效用主义提出了反驳。情况并非如此。罗尔斯式契约论也没有为可接受的非效用主义原则提供支持。

52 康德式契约论

为了达成更优版本的契约论，我们应该回到康德式公式。我们应该问：哪些原则是每一个人——假如这个人知道所有的相关事实，且有力量选择“哪些原则是每个人都会接受的”——都能够合理地意愿的。根据康德式公式，每个人都应当遵循这样的原则，该原则是在上述想象情况下每个人都能够合理地选择的。

53 斯坎伦式契约论

根据斯坎伦的部分类似的公式，每个人都应当遵循无人能够合情理地予以拒绝的原则。既然斯坎伦所诉诸的主张是关于在部分道德的意义上什么是合情理的，看来就可能有人认为：如果我们接受斯坎伦的公式，那也不会对我们的道德思考有任何意义。但情况并非如此。

斯坎伦曾宣称，其公式给出了对不当性本身或者说关于什么是行动之为不当的某种说明。但还是这样说为好：契约论公式描述了使得行动不当的诸种事实之一。斯坎伦的观点现在采取了这种形式。

54 道义式信念的限制

22 在运用任一种契约论公式时，契约论者必须主张，不能诉诸我们关于哪些行为为不当的直觉信念。如果诉诸这样的道义式信念，这些公式就将一无所成。有些契约论者主张，我们应该从不诉诸此类直觉的道义式信念，该信念包含的不过是偏见或文化制约。我们应该拒绝这样的主张。在力图确定哪些行动不当时，我们必须诉诸这样的直觉信念。契约论者应该代之以这样的主张：尽管当我们在弄清契约论公式的蕴含时不应该诉诸这样的信念，但随后在鉴于这些蕴含而决定是否应当接受该公式时，我们可以诉诸这些信念。

第十六章 后果论

55 后果论的理论

无论持有何种道德观点，我们都可以在蕴含不偏不倚理由的意义上使用“最好的”这个词。在此意义上，某种后果如果从不偏不倚的视角看是每个人都最有理由想要的，那就是最好的。应该将这些后果看作是涵盖行为的，并且后果的善性可能部分地取决于与过去相关的事实。后果论的道德理论所诉诸的，仅仅是关于事情如何进展会最好的主张。直接后果论者将该标准运用于一切事。将该标准用于行为的人是行动后果论者。间接后果论将该标准直接地用于一些事，但间接地用于其他事。例如，根据某些动机后果论者（Motive Consequentialists）的观点，尽管最好的动机是那种其被每个人持有将使事情进展最好的动机，但最好的或正当的行为并非会使事情进展最好的行为，而是为拥有最好动机的人所做出的行为。间接后果论可能采取多种其他的形式。

56 后果论的准则

根据准则后果论者的观点，每个人都应当奉行其被每个人奉行将使事情进展最好的准则。按照各种讲得通的或被广泛接受的、关于合理性的观点，康德原版的自然法公式允许一些人成为准则后果论者。

57至62 康德主义论证

根据一个版本的

23

规则后果论：每个人都应当遵循其被普遍接受将使事情进展最好的原则。

我们可将这样的原则称作最优的。

康德主义者可以论证道：

每个人都应当遵循这样的原则，其被普遍接受是每个人都能够合理地意愿或选择的。

每个人都能够合理地选择他们有充足理由去选择的任何东西。

存在某些最优原则。

这些是每个人都有最强的不偏不倚理由去选择的原则。

任何人选择这些原则的不偏不倚理由，都不会决定性地弱于与之相冲突的任何相关理由。

因此

每个人都有充足的理由去选择这些最优原则。

不存在任何其他的、每个人都有充足理由去选择的、明显为非最优的原则。

因此

仅有这些最优原则是每个人都有充足理由去选择的。

因此

每个人都应当遵循这些原则。

24 该论证的第一个前提是康德式契约论公式。这个论证有效，且其他前提为真。因此，这一康德式公式就要求我们遵循上述的规则后果论原则。

我们或许怀疑，这一康德主义论证必定至少有一个后果论前提。如果这一点为真，那么该论证就是无关紧要的。但是，该论证的所有前提均没有假定后果论为真。在无需任何后果论前提的情况之下，该论证如何能有效地蕴含着后果论的结论，其解释如下：

后果论者诉诸的相关主张是：从不偏不倚的视角看，对于每个人来说选择什么将是合理的。对后果论最有力的反驳，来自某些我们关于哪些行动为不当的直觉信念。

契约论诉诸的相关主张是：按照某种使得选择为不偏不倚的方式，对于每个人来说选择什么将是合理的。在契约论的道德推理过程中，我们不能诉诸我们关于哪些行动为不当的直觉信念。

既然两种理论都诉诸对于每个人来说不偏不倚地选择什么将是合理的，而且契约论者要求我们搁置非后果论的道德直觉，那么我们就应该预料到，含有

某种契约论前提的有效论证能够得出某种后果论的结论。

我们可以得出又一结论。我已宣称，存在某些对于康德普遍法则公式的决定性反驳。为了免于这些反驳，康德的公式必须被修正。就其经修正的最佳形式而言，该公式要求我们遵循这样的原则：其被普遍接受是每个人都能够合理地意愿或选择的。根本没有每个人都能够合理地选择的、明显为非最优的原则。因此，除非如我论证过的，每个人都能够合理地选择最优原则这一点为真，否则该公式就不可能成功。除非在其经修正的形式中，康德的普遍法则公式蕴含规则后果论，否则该公式就不可能成功。

第十七章 结论

63 康德式后果论

根据行动后果论的原则，每个人都应当总是做使得事情进展最好的任何事。这并非其被普遍接受将使事情进展最好的原则之一。因此，康德式公式不要求我们成为行动后果论者。 25

根据康德式公式的另一版本，每个人都应当遵循这样的原则，其被普遍遵循或被成功地奉行是每个人都能够合理地意愿或选择的。康德式公式的这一版本，蕴含着一种显然更接近于行动后果论的规则后果论版本。

既然康德式契约论蕴含着规则后果论，那这两个理论就可以合并。要么通过被普遍接受，要么通过被普遍遵循，原则可以成为普遍法则。根据

康德式规则后果论：每个人都应当遵循其为普遍法则将使事情进展最好的原则，因为它们是每个人都能够合理地意愿其为普遍法则的仅有原则。

64 攀登山峰

我们可以论证说，一旦仅有一组原则是每个人都能够合理地意愿其为普遍法则的原则，那它们就是无人能够合情理地拒绝的仅有原则。如果这一点正确，那么上述被合并的理论就可以同时包含斯坎伦的公式。根据我们可称之为的

三重理论：某一行动是不当的，恰好当这样的行动为某原则所不容，该原则是某种最优的、唯一地普遍可意愿的，以及不可合情理地予以拒绝的原则。

26 我们如果接受这一理论，就应该承认，行为可能有其他的致不当特性。三重理论应该宣称，其所描述的是单一而复杂的高阶特性，它可以纳入所有其他的致不当特性。该理论如果成功，就会描述这些其他特性的共性。

该理论可能取得成功，因为它具有许多有道理的蕴含。康德式公式与斯坎伦式公式本身也是言之成理的。在三重理论的三种成分之中，规则后果论在一定意义上是最难得到辩护的。有些规则后果论者诉诸如下主张

(Q) 最终要紧的仅在于事情进展如何。

这一主张本身很有道理。如果我们拒绝(Q)，那是因为该主张支持行动后果论，而行动后果论与我们关于哪些行动为不当的直觉信念之间的冲突过于频繁或过于强烈。规则后果论与这些直觉信念之间的冲突则要少得多，弱得多。但如果规则后果论诉诸(Q)，那么其观点将面临某种强有力的反驳。按照这种观点，尽管我们知道我们的行动会使事情进展最好，但该行动为最优原则所禁止，因而这么做依然是不当的。我们可以有理由反对说：如果最终要紧的仅在于事情进展如何，那么这样的行动就不可能是不当的。

康德式规则后果论避免了上述反驳。按照该观点，作为基础的，并非关于最终要紧之事的信念，而在于如下信念：我们应当遵循“每个人都能够合理地意愿其为普遍法则”的原则。

在怀疑存在道德真理的理由之中，最强的理由之一来自某些类别的道德分歧。如果我们和他人持有相冲突的观点，而我们毫无理由地相信，我们是更可能正确的人，那就应该至少使我们怀疑自己的观点。它或许还给予我们以理由去怀疑，我们是否有任何人可能正确。

一直以来人们广泛相信，康德主义者、契约论者，以及后果论者之间存在着这样的深刻分歧。我已论证，情况并非如此。这些人是在不同的侧面攀登同一座山峰。

附录A 状态给予的理由

27 当某些事实使得我们如果持有某一信念就会好些，这些事实就给了我们对象给予的理由去希望持有该信念，并尽可能使自己持有之。如下的补充毫无意义：我们也有状态给予的理由去持有该信念。尽管我们无法立刻回应这些所谓的理由，但我们的心理可能发生变化。在相信我们如果拥有某种认识上不合理的信念就会好些时，我们可能以某种非自愿的直接方式不知不觉地开始持有该信念。但这不应被看作是对状态给予的理由做出的回应。这样的理由也绝不可

能与我们的认知理由相冲突。看上去更有道理的主张是：当某些事实使得我们如果拥有某欲望就会好些，这些事实就给了我们拥有该欲望的某种理由。但我们也有拒绝这一主张的强理由。

附录B 合理的不合理性与高契尔的理论

高契尔主张：当我们已合理地致使自己拥有某一习性时，依此习性行动对我们来说就是合理的。这个主张有若干讲不通的蕴含。致使我们自己相信按照这样的习性行动是合理的，这尽管可能是合理的，但这一事实并不能说明该信念为真。高契尔还主张，如果我们接受契约论的一种霍布斯式版本与某种最低度道德，那么他的论证就揭示出，我们被合理地要求绝不做出不当之举。由于这一论证失败，我们就毫无接受高契尔的最低度道德之理由。

附录C 道义式理由

在为规则后果论的康德主义论证之前提（E）辩护的过程中，我提出

（X）如果最优原则要求某些我们认为是不当的行动，那么我们就不会有决定性的非道义式理由这么做。任何这样的决定性理由都必须²⁸是道义的，其意思是这些理由是由这些行动的不当性提供的。

某些人在宣称某一行动不当的时候，其意思是：我们有决定性的道德理由不这么做。这些人会否认存在任何道义式理由。按照这一观点

（2）“某个行动是不当的”这一事实是二阶事实，即：其他某些事实给予我们决定性的道德理由不这么做，而“我们拥有这些理由”这个事实不会给予我们进一步的理由不这样做。

如果（2）正确，那（X）就会被部分地动摇。然而，鉴于我们大多数人用“不当的”所表达的意思，我们可以有正当理由拒绝（2）。而且正是在最有必要运用（X）的那些情形之中，（2）看来最讲不通。

目 录

导论	vii
序言	xviii
内容提要	xxxix

第一部分 理由

第一章 规范性概念	3
1 规范性理由	3
2 涉及理由的善性	8
第二章 客观理论	13
3 两类理论	13
4 回应理由	16
5 状态给予的理由	18
6 享乐式理由	20
7 不合理的偏好	23
第三章 主观理论	25
8 理由的主观理论	25
9 人们为什么接受主观主义	31
10 分析性主观主义	35
11 痛苦论证	38

第四章	进一步的论证	46
	12 全部或全不论证	46
	13 不融贯性论证	53
	14 理由、动机和福祉	62
	15 支持主观主义的论证	67
第五章	合理性	71
	16 实践合理性与认知合理性	71
	17 有关理由的信念	77
	18 有关合理性的其他观点	82
第六章	道德	87
	19 西季威克的二元论	87
	20 最深刻的问题	96
第七章	道德概念	103
	21 无知或错误信念下的行动	103
	22 其他类型的不当	116

第二部分 各种原则

第八章	可能的同意	129
	23 强制与欺骗	129
	24 同意原则	131
	25 给予同意的理由	134
	26 一个多余的原则?	139
	27 实际同意	141
	28 道义式信念	148
	29 极端要求	154
第九章	纯粹作为手段	158
	30 纯粹手段原则	158

31 作为手段与纯粹作为手段·····	165
32 作为手段的伤害·····	171
第十章 尊重与价值·····	175
33 对人的尊重·····	175
34 两类价值·····	177
35 康德式尊严·····	180
36 正当与善·····	184
37 促进善·····	189
第十一章 自由意志与应得·····	197
38 道德所要求的自由·····	197
39 为什么我们不可能应得痛苦·····	201

第三部分 理论

第十二章 普遍法则·····	213
40 不可能性公式·····	213
41 自然法公式与道德信念公式·····	221
42 行动者的准则·····	224
第十三章 如果每个人都那么做会怎样? ·····	235
43 群己困境·····	235
44 门槛反驳·····	241
45 理想世界反驳·····	244
第十四章 不偏不倚·····	252
46 黄金规则·····	252
47 稀罕性反驳与大赌注反驳·····	260
48 不可逆反驳·····	262
49 一个康德主义的解决方案·····	266

第十五章 契约论·····	270
50 合理的协议公式·····	270
51 罗尔斯式契约论·····	272
52 康德式契约论·····	280
53 斯坎伦式契约论·····	284
54 道义式信念的限制·····	289
第十六章 后果论·····	294
55 后果论的理论·····	294
56 后果论的准则·····	298
57 康德主义论证·····	300
58 自利的理由·····	302
59 利他理由与道义式理由·····	306
60 致不当特征的反驳·····	309
61 决定性的非道义式理由·····	314
62 每个人都能够合理地意愿什么·····	317
第十七章 结论·····	322
63 康德式后果论·····	322
64 攀登山峰·····	328
附录A 状态给予的理由·····	336
附录B 合理的不合理性与高契尔的理论·····	347
附录C 道义式理由·····	360
译后记·····	364
尾注·····	367
引注·····	399
参考文献·····	418
索引·····	425

第一部分
理 由

第一章 规范性概念

1 规范性理由

我们是既能理解也能回应理由的动物。这些能力（abilities）给了我们伟大的知识和力量（power）来控制我们在地球上的未来生活。尽管其他地方也可能有生命，但可能没有像我们这样的生命，我们也许是整个宇宙中绝无仅有的理性存在者。

31

我们能理由去相信某些东西、做某些事、持有某些欲望和目标、具有如害怕、遗憾和希望等诸多态度和情绪。理由是由事实给予的，诸如“某人的手指正放在扳机上”，“打电话给救护车可以救命”等这样的事实。

我们很难解释，理由这个概念或“一个理由”这个短语意指什么。我们可能会说，若事实“算作支持”（count in favor of）我们具有某种态度或以某种方式行动，则它们就给予我们理由。但是“算作支持”的大意就是“给予……理由”。就像某些其他的根本性概念，如涉及我们思维中有关时间、意识和可能性等的概念一样，理由概念也是无法定义的，也就是它无法仅仅通过使用语词得到有用的解释。我们必须以不同的方式来解释这种概念，让人们去思考使用这些概念的那些思想。这样的思想之一例是：我们总是有理由想要避免处于痛苦之中。

我会说，我们还能有并未意识到的理由。假定我问医生，“我对苹果过敏，由此是否有理由不吃别的某种食物？”如果医生知道胡桃会致我于死，其答案就是“有”。这个事实给了我理由。

有些人说，不是事实给予我们理由，而是这些事实是对于我们的理由（reasons for us）。有些人则说，要有某种理由，我们必须意识到给予我们

这个理由的那个事实。不过，这些人的主张与我的主张并不冲突，因为这些只不过是不同的方式来说同一件事。那个医生也许会说，“没有，你没有任何理由不吃任何其他的食物。但在我告诉你吃胡桃会杀死你后，你就有理由不吃它。”我已经有这个理由是个更简单的说法。

当必须在多种可能的行动之中进行选择之时，我们的多种理由之间有可能发生冲突，它们在我们所称的力量、强度或权重方面可能有别。如果喜欢胡桃，这个事实给予我理由要吃它；但是，如果我吃它会死，这个事实就给了我一个比前一理由更强或更有份量的冲突理由不去吃它。如果我们有多种理由以某种方式行动，这些理由加起来就可能强过或压倒某种单个看来更强的冲突理由。如果我能够要么使你免于10小时的痛苦，要么使你免于9小时的痛苦，同时使另一人免于8小时的痛苦，那么我就有一组更强的理由以第二种方式行动。正如我们更简洁的说法，我们更有理由（more reason）这样做。

我们以某种方式行动的理由如果比以其他任何方式行动的理由都要强，那么这些理由就是决定性的，以这种方式行动就是我们最有理由（most reason）做的事情。如果这些理由比相反的理由要强得多，我们就可以称它们为强决定性的。尽管大多数类型的理由都只在某些情形下是决定性的，但有些类型的理由可能总是决定性的。例如，按照某些观点，我们总是有决定性的理由不做不当行动。

当我们意识到某些事实给予我们决定性理由要以某种方式行动时，如果对这些理由的意识使得我们去做或努力去做我们秉持这些理由会做之事，那么我们就是在回应这些理由。正如忽略某人的求救就是没有回应求救一样，如果我们忽略这些理由，我们就没有回应它们。

我们常常没有决定性的理由或最有理由去做任何事情，因为我们具有充分的理由或足够的理由去做两种或更多事情。如果做某事的理由不弱于去做其他任何事情的理由或被其所压倒，那么该理由就是充分的。例如，我们也许有理由去吃桃子、李子或梨子，或者选择法律或医学职业，或者把我们的收入捐给Oxfam或其他类似的援助机构，诸如Mèdecins Sans Frontières。当两个相冲突的理由没有更强者的时候，这极少是因为它们的强度恰当相等。尽管存在有关不同理由的相对强度之事实，但这些事实通常是不精确的。

理由之间的关系可能更为复杂。例如，有些事实给予我们理由去忽略某些其他的理由。如果我要判断谁应得某种奖，这就给予我理由忽略某个竞争者是我最好的朋友这个事实。有些事实并不能在所有情形中都给予我们理由，而只

有组合其他某些事实才给予我们理由。我主要讨论更简单的理由。

当我们有决定性理由或最有理由进行某个行动时，这个行动就是在我们蕴含决定性理由的意义上“应该”或“应当”（should or ought）去做的。即使我们从来不使用“决定性理由”或“最有理由”这样的短语，我们绝大部分人还是在蕴含这些理由的意义上使用“应该”或“应当”。“必须”也有同样的意义。这些语词蕴含不同强度的理由。我可能会说你应该看某个电影，你应当戒烟，你“绝不能”（mustn't）碰正通电的电缆。尽管“应该”这个语词用得更多，而“必须”这个词更有力量，但我还是使用更为清楚的“应当”。

我们除了可以问在蕴含决定性理由的意义上我们应当做什么，还能问我们应当理性地做什么。我们如果是在非技术性的日常意义上称某个行动是“合理的”，那也可以用“明智的”“合情理的”“伶俐的”和“聪明的”来表达这类表扬与赞同。我们也用“不合理的”来表达可以用“不明智的”“愚蠢的”和“白痴”以及“发疯”等表述的那类批评。要表达较弱的批评，我们可以使用“不够理性的”。

当我们必须在几种可能的行动中做选择时，可能会有几种事实给予我们理由以这些方式行动。我称它们为给予理由的相关事实。我们应当理性地去做什么，在部分程度上取决于我们有关这些事实的信念。这些信念可能包含一些我们尚未明确意识到的假设——诸如我们吃胡桃、接触电缆或推开旋转门不会伤害自己和他人。如果我们具有相关事实的某些信念，且这些信念如为真就会给我们以某种方式行动的理由，我就将此称作其真实性会给予我们理由的信念。我认为，在大多数情形下我们的某种可能行动会是

34

合理的，如果我们具有这种相关事实的信念，且其真实性会给予我们充分的理由做这种行动。

我们应当理性地去做的，如果这些理由是决定性的。

不够合理的，如果我们具有的信念为真会给我们清楚且决定性的理由不做这种行动。和

不合理的，如果这些理由是强决定性的。

按照这种观点，一旦知道所有的相关事实，我们应当理性地去做的，就等同于我们在蕴含决定性理由的意义上应当做的。但是当我们无知或具有错误信念时，这些应当就可能会发生冲突。假定你在沙漠行走时激怒了一条毒蛇。你

认为要活命就必须逃跑。事实上，你必须静止不动，因为这条蛇只攻击移动的目标。鉴于你的错误信念，你静止就是不理性的，你在理性上应当逃跑。但是在“蕴含决定性理由”的意义上这不是你应当做的。你没有任何理由要逃跑，而且有决定性的理由不要逃跑。既然静止是唯一能活命的行动，你就应当静止。

35 有人会说：你的确有逃跑的理由，该理由来自你逃跑会救你的命这个错误信念。如果说错误的信念能给予人们理由，那么我们还需要补充一句：这些理由并不具有任何规范性力量，也就是它们并不算作支持任何行动。这样的理由是我们在决定个人最有理由做什么之时所必须忽略的。我们最好是以其他的方式来描述这种情形。我们一旦具有这样的信念——其真实性能给予我们以某种方式行动的理由，那就具有了我所称的、依此方式行动的显见理由。如果这些信念确实为真，这种显见理由也是真正的理由。如果这些信念是错误的，我们就具有仅仅是貌似理由。在愤怒的蛇的情形中，既然你具有“逃跑才能救你命”这个错误信念，那么你仅仅有显见的理由逃跑。当我们相信自己具有某种理由时，我们具有了一类不同的显见理由。我们现在可以主张所有理由都具有规范性力量。当我们给人们提建议时，我们可以忽略由于人们的错误信念而给予的、仅为显见的理由。但是人们要做的合理的事情取决于他们的显见理由，无论这些理由是真正或仅仅是显见的。

我们现在可以从可能的行动转向实际的行动。我认为，在绝大部分情形中，我们的行动是

合理的，如果我们做某种行动是因为我们具有相关事实的信念，而这些信念为真会给予我们充分的理由做这种行动；

和

不合理的，如果我们做某种行动，我们具有的信念若为真会给我们清楚且决定性的理由不做这种行动。

如果这些信念是明确地意识到的，则这种行动将是最不合理的。若理由不是那么清晰，或者只是弱意义上决定性的，则我们的行动就只是不够合理的。例如，明明知道我们有可能会因抽烟上瘾且短命，还要抽烟就是不合理的。而买一些我们不大可能会读的书或者去拨打电话来报告我们的电话坏了，这只是

不够合理的。

尽管正是事实给了我们理由，但行动的合理性为什么反而要取决于信念，这个值得解释一下。当我们试图决定自己或他人应当做什么时，重要的是给予理由的事实。在激怒的蛇的例子中，你应当静止不动，因为这是你活命的唯一方法。当我们问某个人的行动是否合理时，我们的目标不同。我们问的是这个人是否该受到诸如“愚蠢”、“笨”或“疯狂”之类的批评。在不知情或具有错误信念时，人们可能做在“蕴含决定性理由”的意义上他们不应当做的事情。但这些人不该受批评，因为他们的错误信念若为真，则会给予他们充分的理由这么做。至少在大部分情形下，这足以使得他们的行动是合理的。如果你由于错误地相信逃跑能救命而逃跑，那么你的致命行动不会是“愚蠢的”“笨的”或“疯狂的”。你可能仅仅是运气不好。

许多人主张，要理性地行动，我们只依据那种若为真就会给予我们充分理由做某种行动的信念去做是不够的。仅当我们的信念是合理的，我们的行动才是合理的。我后面将论证，这并不是最好的看法。

要想成为完全理性的，我们可能还需要满足某些其他的合理性要求：避免某种不一致性以及避免在我们的意图、信念和其他心理状态之间的不匹配等。例如，可以合理地要求人们不持有矛盾的意图，并且打算做我们应当做的事情。尽管这些要求引出了几个有趣的问题，但我在此将不予讨论。我认为，有关理由的问题是更为根本的。尽管我们是否想要我们有理由想要的事情，是否做我们该做的事情常常事关重大，但它对于我们是否不一致且由此而没满足合理性的要求来说，一般是无关紧要的，至少不是很重要。有些人则主张，要成为理性的，我们没必要对理由或显见理由做出回应，我们只需要满足这些合理性要求就够了。我将在后面给出反对这种观点的几个论证。

我将简要地提及其他几个类似的问题，然后把它们搁置一边。在要决定做什么而未能完全了解所有的相关事实之时，我们必须将决策基于我们的信念以及现有的证据。在这种情形下，我们可以这样问：在我们所称的证据相关的意义上我们应该或应当做什么。看起来在这样的情形下，我们应当努力去做我们最有理由做的事情。但这种努力可能风险太大，或者不大可能成功。我们通常应当按照其目标定得不是很高而更可能达成的方式行事。如果许多人生命攸关，我们该做的就是那种肯定能救大多数人的行动，而不是做那种能救所有人但成功概率很小的行动。

我们在有风险或不确定性的情形下应当做什么，这在实践上特别重要。许

多哲学家、决策理论家还有其他人都对这些问题做了很多讨论。然而，尽管有关理由的其他某些问题更为根本，但甚少得到讨论，这些也是人们分歧更深的问题。既然我将主要探讨这样的问题，我将主要考察如下情形：其中我们知道给予理由的所有相关事实。

这些主张一直都是有关规范性理由的。当我们具有这样的理由或显见理由且出于这种理由而行动时，这种理由就变成了驱动性（**motivating**）理由。例如，如果我不吃胡桃，我的驱动性理由就是，正如医生告诉我的那样，吃它们会要我的命。当我们做某种行动只有驱动性理由时，这个区分就最为清楚。如果你逃离愤怒的蛇，你的驱动性理由就是由你错误的信念（这样做会救你的命）所提供的。但正如我说过的，你没有任何规范性理由要逃跑。你仅仅是认为你有。在一个不同的例子中，我们也许会声称：“他的理由是想报复，但那根本不是那样做的理由。”既然我不拟讨论人们行为的原因，就将鲜少讨论驱动性理由。

除了问我们在蕴含决定性理由的意义上该做什么，该理性地做什么，我们有时候还会问：在其他几种道德意义上，我们应当做什么。其中大多数意义至少与蕴含决定性理由的意义在两个方面有所不同。首先，我们通常具有的决定性理由不是道德理由。例如，如果我想赶上火车，我可能有决定性的理由现在离开会议。如果我憎恨挤车，那么我最有理由住在工作地点的附近。这些可能都不是我在道德上应当做的事情。其次，当我们相信自己在道德上该做某事时，我们可能并不认为自己有决定性的理由这样做。按照某些观点，我们根本没有任何理由去做道德上该做的事情。在这些章节中，我将首先讨论理由，在此之后才转向道德。

我们很容易把“蕴含决定性理由意义上的‘应当’”、“理性上的应当”和“道德上的应当”混淆在一起。因此，我不讨论我们在蕴含决定性理由意义上应当做什么，而是讨论我们具有决定性理由或最有理由做什么。

2 涉及理由的善性（**goodness**）

我们接下来会讨论事物可好可坏的一些方式。当某物

在我们称为蕴含理由的意义上为好的时候，我们大略是指有关此物的本质或属性的某些事实某些情境下给予我们或他人很强的理由

做出正面回应，诸如想要、选择、使用、生产或保存此物。

例如，有些书是好的，它让人愉悦、鼓舞人心或包含有用的信息。有些药物是最好的，它最安全且最有效。这些事实会给我们或他人理由来看这本书或吃这种药。“更好的”、“坏的”、“更坏的”或“最坏的”也有类似的意义。

事物也可在其他意义上是好的或坏的。例如，如果我声称某棵树的根是好的，鼯鼠具有坏的视力，或者最好的比喻是：

男管家整个上半身就像冰冻了一样。^① (ice formed on the butler's upper slopes.)

以及最好的回文不是“Madam I'm Adam”，而是
A MAN A PLAN A CANAL: PANAMA.

我的意思并不是说“好的”、“坏的”和“最好的”的这些用法都是蕴含理由的。鼯鼠没有理由去带眼镜，我们也没理由被“男管家的背部结了冰”所娱乐。“好的”的许多用法只是意指某物满足某个标准。但我认为，“好的”和“坏的”的最重要的用法是“蕴含理由”的用法。

39

斯坎伦主张，当某物在这个意义上是好的的时候，此物的善性（goodness）不可能给予我们理由。这种善性是“具有‘能够给予我们某些理由’的其他属性”的属性，“我们有这些理由”这个二阶事实本身不会给我们任何理由不这么做。

我认为这种观点需要做一个小小的修正。如果某种药或书是最好的，那就的确能主张：该事实给予我们吃这个药或看这本书的理由。但是，这些并不是进一步的独立理由，它们是派生的，因为其规范力量完全来自“使得该药或书是最好的”的那些事实。正是由于这一点，如下的主张就很古怪：我们具有服用某药的三个理由，即该药是最安全的、最有效的以及最好的。既然这种派生理由没有任何独立的规范力量，在这样的主张中提及它们将是误导性的。

我们的许多行动理由是由这样的事实提供的，它们会

^① 这是指男管家不赞同某人所说的话，用其面部和上半身的变化来显示这一点。

对我们是好的，意思是对我们是有利的、有益于我们或有助于我们的福祉。

40 当人们说某事对我们是好的或有益于我们时，他们通常指的是这件事对我们的健康、个性或收支平衡有好的效果。在我意指的更宽泛的意义上，某事如果是构成我们福祉的生活特征之一，那么它就对我们是内在地好的或自身就是好的，因为正是这些特征使得我们的生活值得过。如果某事具有对我们内在地好的效果，那么它就是工具性地好的。基于享乐主义理论，我们的福祉大致在于快乐和幸福，以及免于痛苦。基于诉诸其他实质善（substantive good）的理论，我们的福祉在部分程度上还在于其他某些状态或活动，比如爱与被爱、道德上善、行为得体以及各种各样的其他成就。按照基于欲望的理论，我们的福祉在于满足我们的某些欲望，比如我们对自己生活的知情欲望。基于任何可行的理论，享乐主义至少涵盖了大部分的事实，因此我的例子将常常涉及享乐式福祉。

我们有自利的理由关注自己的福祉，但还有利他的理由关注他人的福祉。这些是为了自己或他人的缘故而想要某事发生的理由。“自利的”并不意指“自私的”。因为最无私的人也有自利的理由，他们也要关注自己未来的利益。

我们有强有力的理由关注其他某些人的利益，诸如我们的近亲和爱人。像自利的理由一样，这些利他的理由是

个人相对的或偏向性的，意思是这些理由特别关注那些与我们具有某种关系的人的福祉。

我认为，我们也有理由关注每个人的福祉，这种理由是

不偏不倚的，意思是

(1) 这些是关注任何人之福祉的理由，而不管其与我们的关系如何，

因此

(2) 即使我们的情境给予我们不偏不倚的视角^①，我们还是具有这样的理由。

我是在其本义上来使用“视角”这一术语，不同于我们在谈论理由时从金钱、审美以及诸如此类的其他视角所使用的那种松散意义。当我们考虑的可能事件影响或涉及的全都是陌生人时，我们具有一种不偏不倚的视角。当我们实际的视角不是不偏不倚时，我们能从一种虚构的不偏不倚的视角来思考可能的事件。做法是想象在相关方面类似的可能事件，除了其中所涉及的全都是陌生人。

41

我认为，我们具有不偏不倚的理由要同等地照顾所有人的福祉。这是一个实质性的信念，不是由我对不偏不倚理由的定义所蕴含的。按照广泛持有的其他某种观点，我们具有不偏不倚的理由更多地关照某些人，诸如那些道德上的好人，或那些最有能力的人。当我们的视角是不偏不倚的时候，那并不确保我们是不偏不倚的。我们可以更多地关照某些陌生人的福祉，诸如那些与我们更相像的人，或者我们喜欢其长相的人。但我认为，我们并没有任何理由更多地关照这些人的福祉。

我们接下来描述事件可能好或坏的两种方式。当我们称某个可能事件

在蕴含理由的意义上对某人是好的时候，我们指的是存在某种事实给予此人自利的理由想要这个事件发生，给予其他人利他的理由，为了此人的缘故想要或希望这个事件发生。

这个定义看来所说甚少，因为它参考了自利的理由。然而，正如我们所见，我们是否有这种理由是有争议的。

我们称两个事件之一

在蕴含不偏不倚理由的意义上更好时，我们指的是从一种不偏不倚的视角看，每个人都有更强的理由想要这个事件发生，或希望它会发生。

① 尽管根据上下文，这里“不偏不倚的视角”改为“偏向性视角”既是对的，也更好理解，但帕菲特在给笔者的信中强调，他在此针对的是这样一些人：他们接受了基于欲望的主观主义理由观，从而认为，即使我们具有的是完全不偏不倚的视角，我们依然没有理由去关爱某些事物。

我认为，如果瘟疫或地震杀死更少的人，如果人或动物不再处于疼痛之中，这会是更好的。这类好处（goodness）是非个人性的，意思是，当我们称事件在这种意义上是好的，我们并不意指这个事件会对某个人或某个群体是好的。但许多事件之所以是非个人性地好的，是因为它们对某个人或更多人是好的。对这些人有益就是使得这些事件是非个人性地好的因素。既然每个人都有理由想要这种事件发生，这种非个人性好处就包含了全部人（omnipersonal）的理由。

如果某个可能事件在这种意义上对某个人是好的，或非个人性地好的，我们就可以主张，这个事实真正地给予我们理由想要这个事件发生。但正如此前，这个理由是派生的，因为这个理由的力量派生自“使得这个事件对于个人是好的或非个人性地好的”那个事实。当我们在这种意义上使用“对……是好的”和“好的”，这些不过是更简洁的方法来蕴含存在着赋予理由的其他事实。不像一个理由这种概念以及蕴含决定性理由的应该或应当这些概念，这些版本的“好的”这个概念并不是根本性的。

基于某些被广泛接受的理由观，没有事件可以在这些意义上是好的或坏的，或者是非个人性地好的或坏的。如果这种观点是对的，这将极大地影响到我们最有理由想要或去做什么。但我将论证，我们应当拒绝这些观点。

第二章 客观理论

3 两类理论

43

“欲望”一词通常指向我们的感官欲望或嗜欲（*appetite*）、受到某事的吸引和发现某物有吸引力这个想法。我会在广义上使用“欲望”，它指涉受到驱动、尽可能想要某事发生或在某种程度上倾向于让它发生的任何状态。“想要”这个词已经具有两种意义。如果你和我正计划如何一起过几天，我可以毫不自相矛盾地说，“我想要我们去做的事情，不是我想我们要去做的，而是你想我们要去做的。我在广义上想要的不是我在狭义上想要的，而是你在狭义上想要的。我想要我们去做的是你受吸引去做的，或者你觉得有吸引力的事情，即使我并不觉得它有吸引力。

有些人认为：“只要人们是自愿行动的，他们所做的就是他们想做的。做我们想要的事情是自私的，因此每个人都是自私的。”心理利己主义的这种论证是失败的，因为它起初使用的是广义的想要，后来使用的是狭义的想要。如果我自愿牺牲自己救几个陌生人的命，则尽管我做的事情是我在广义上想做的事情，但我的行动不是自私的。

我们的欲望具有对象，它就是我们想要的东西。这些对象是在包括行动、过程和事态意义上的一切事件。我们可以正确地说想要其他类型的事物。我可能想要在威尼斯有个公寓、一杯水或一个钢琴老师。有些逃犯正是警察想要的。但我们真正想要的是：拥有某些东西、住在某个地方、喝点东西、接受某人教育、发现或使用某些东西或者与人或物处于某种关系之中。我不会说我们想要某事发生，而是更简洁地说我们想要这样的事。

如果我们将某事作为目的或为了此事自身之故而想要它，那么我们的欲望

就是目的论的或目的式的（teleological or telic）。如果我们想要以某事作为手段，因为该事将要或可能导致我们想要的其他某事，则我们的欲望是工具性的。我们想要某个行动或事件，这种想要既可以作为目的，也可以作为达到其他目的的手段。兴高采烈地追求重大真理和想要孩子时的做爱就是两个这样的例子。一旦决定要努力满足某种目的式欲望，我们就由此使这种欲望的满足成为我们的目标之一。

我们通常拥有长长的工具性欲望之链，不过，这种链条总是会从某个目的式欲望开始，或者以其为根据。例如，我要治病，但不是为了治病本身之故，只是想要恢复健康；而我想要健康只是想要完成某部伟大的小说；而想要完成这部小说只是为了身后之名。最后的欲望依然可能是纯粹工具性的，因为我想要这种名声也许只是想驳斥我的批评者，或者只是为了增加后代的收入。但如果我想要身后之名是为其自身之故，那么这个目的式欲望将启动这个特定的链条。

心理享乐主义者主张，在所有这种工具性欲望链条之始都有某种趋乐避苦的目的式欲望。这是错的。有些人持有这种观点是把它混淆为这种观点：我们想到这种欲望的实现预先就会感到快乐，而想到它实现不了就会感到痛苦。这也是错的。即使它是对的，它也无法表明我们真正想要的总是趋乐避苦。例如，若我想要身后之名，一想到我仙去之后，人们会如何铭记我，赞美我小说的伟大，我就会快乐。但这并不表明我是为了快乐而想要这种名声。相反，这种快乐依赖于我们是为了这种名声自身的缘故而想要它。另一例是这个事实：既然仅当我们想赢时才能享受比赛，因此只是想要享受比赛是不够的。

除了还想要这样的东西之外，有些人甚至不把快乐作为目的。假定我们认识一个百折不挠且雄心勃勃的政治家，我们看到他正晒着太阳，喝着香槟。当我们问他在干什么时，他的回应是“享受生活”。由于我们知道这个人的个性，这个回应是不可理解的。这个人从不做只为了享受的事情。然后他解释，医生警告他，除非他进行这种享受，否则他的健康情况会恶化，由此阻止他追求权力。我们的困惑消失了。这个人想要这种快乐，不是为其自身的缘故，只是作为一种手段。

对于我称为的实践理由主要有两类观点。按照一类观点，存在某种事实既给予我们理由要有某种欲望和目标，也给予我们理由去尽可能实现它们。与这些欲望或目标对象有关的事实，或者与我们想要实现或尽力实现之物有关的事实，会给予我们这种理由。由此我们可以称这种理由为对象给予的。如果我们相信所有实践理由都是这一类的，我们就是理由的客观主义者，这种人接受或

假定理由的某种客观理论。

对象给予的理由之提供者是这样的事实，它使得某些结果值得产生或防止，使得某些事情值得为了自身的缘故而去做。在大多数情形下，这些给予理由的事实也使得这些结果或行动对特定的人是好的或坏的，或者是非个人性地好的或坏的。这样我们也能称这些客观理由理论是基于价值的。

按照另一类观点，我们的行动理由是这样的事实所提供，或者是依赖于它们：什么会满足我们当下的欲望或目标。有些这样的理论诉诸我们实际的欲望或目标。其他理论则诉诸在仔细地考虑一切相关事实后我们会有的欲望或目标，或者会做出的选择。由于这些都是有关我们的事实，我们可以称这些理由为主体给予的。如果我们相信所有实践理由都是这一类的，我们就是理由的主观主义者，接受理由的某种主观理论。

这两类理论是极为不同的。按照客观主义观点，尽管可以声称，许多行动理由是由“某个行动会达成我们的一个目标”这个事实而给予的，但这些理由是通过这样的事实而获得其力量的：该事实给予我们持有这些目标的理由，正是这些事实使得这些目标在相关意义上是好的或值得追求的。

按照主观主义观点，我们没有任何这样的理由拥有我们的目标。有些主观主义者甚至主张，正是我们用自己的欲望或选择使得事情是好的。例如，科丝嘉在捍卫这样一种观点时写道：

大多数事情是好的是因为人们对它们感兴趣……客观主义者搞反了这种关系……客观主义者不认为，某物是因我们感兴趣而是好的；而是认为，善性存在于对象之中，从而我们应当对它们感兴趣。

这种善性给予我们理由的方式就像是太阳发光的方式那样，“因为它就在那，闪闪发光”。如果主观主义者是对的，那我们就必定是在黑暗中做选择。

主观主义者与客观主义者通常在部分程度上是一致的。按照所有可行的客观理论来说，我们有理由促进自己未来的福祉。既然我们大多数人想促进未来的福祉，主观理论也蕴含着，我们大多数人都有理由这样来做。既然我们想要的事物通常也是值得拥有或获得的事物，我们大多数人持有的许多欲望因此是“两种理论都告诉我们要尽力实现的”。

尽管两类理论通常都赞同我们有理由去实现当下的欲望与目标，但我们具有更强的理由去满足何种欲望，这些理论对此往往意见不一。基于许多主观理

论，理由强度取决于我们欲望或偏好的强度。基于客观理论，理由的强度反而是取决于满足这些欲望会有多好或有多大价值。我们许多人对于价值不怎么大的事物，常常有更强的欲望。许多这样的欲望都涉及对时间的态度，我们往往称之为偏向眼前。我们都偏好很快就会具有的享受体验，即使我们明知，若等一等，会有更大的享受。我们都偏好推迟乏味的苦差事或者不可避免的煎熬，尽管我们明知这种推迟只会使事情更糟。我们都喜欢做“月光族”，尽管我们知道有些钱花在后面会带来更多的利益。在满足这些欲望和偏好时，许多人使得自己的生活更为糟糕。在这些以及其他许多方面，主观理论和客观理论对我们最有理由做什么意见不一。

两种理论之间还有着其他更深刻的分歧。正如我们会看到的，其中一类理论都蕴含着我们具有决定性的理由要做某些事情，而另一类理论则蕴含着我们根本没任何理由要做这些事情，且有决定性的理由不这样做。我们要具有何种欲望和目标，对此这两类理论的理由是完全不同的。

我将论证，我们应当接受某种基于价值的客观理论。按照这些理论，行动理由的力量完全派生于给予我们理由要有这些欲望和目标的事实，后面的理由才是更根本的。

4 回应理由

同样的事实既可以给予我们理由来想要某些事情发生，还可以给予理由让我们以某种方式使这些事情发生。正因为如此，我将理由的这两种类型都称为实践的。尽管这两类理由紧密相关，但我们对它们做出回应的方式却截然不同。我们一旦认识到某些事实给予我们理由要以某种方式行动，通常会这样行动来回应这些理由。这种回应是自愿的，意思是如果我们本不想这样做，我们就可以选择不这样做。但一旦我们认识到事实给予我们强有力的理由要有某种欲望，我们的回应就很少是自愿的。情况很少是这样的：如果我们本不想要这些欲望，我们就可选择不要有这些欲望。例如，我们很少能够选择是否想活着和免于巨痛。即使某个反复无常的暴君威胁要杀我，除非自即刻起我想要被杀，我也还是无法选择去拥有被杀这个欲望。

这同样适用于我们要某种具体信念的认知理由。给予我们这种理由的是与某种信念为真相关联的事实：这个信念为真的证据、逻辑上或以其他方式蕴含这个信念。例如，我们若看到了暗灰色的云层，就有理由相信很快会下雨。

若我们知道金比铅重，而铅又比铁重，这就给了我们决定性的理由相信金比铁重。若我们意识到事实给了我们决定性的理由拥有某种信念，我们对这种理由的回应就是变得持有且一直持有该信念。但我们对这种理由的回应极少是自愿的。我们若有决定性的理由要相信某事，就很少能够选择不相信。如果上面虚拟的那个暴君威胁我，除非自即刻起我不再相信“ $2+2=4$ ”，否则就杀死我，那么我还是无法选择不相信“ $2+2=4$ ”。

有些作者主张，若我们以这种直接且非自愿的方式持有某个信念或欲望，这是我们所做的一个行动或一件事。但我会更狭窄的意义上使用“行动”或“做事”，只用它们来指自愿的行动。许多这样的行动是纯粹心理上的。例如，当你自问是否有足够的时间赶上火车时，你会主动去做心理计算来回答。在这个复杂的行动中，你意图产生这样的结果：你会得到有关此问题的某个信念。但若这个计算让你相信你的时间不够，这个特定信念的获得就不是一个行动，或不是自愿的。例如，你无法选择相信：你能在十分钟内跑十英里，从而赶上火车。

尽管我们若有理由持有特定的信念和欲望，对此很少能够选择如何回应，但我们的回应并不像自动的膝跳反射或潜在香蕉皮上那样，仅仅是发生在我们身上的事件。我们是理性的，部分程度上就在于我们以这种非自愿的方式对这种理由或显见理由做出回应。人们能够问我们为什么会相信某事或想要某物，而我们对此通常能给出理由。

值得追问的是，我们对此种理由的回应是否可以采取其他的形式：总是或通常是自愿的。假定当你认识到某种事实——它给予你决定性的认知理由具有某种信念——之时，你未能通过持有该信念而对这一理由做出非自愿的理性回应。尽管你在旅馆楼梯上看到浓烟大火向你冲来，但你并不相信你的生命有危险。你能纠正错误，选择持有该信念吗？

答案很可能是不能。首先假定，正如你不相信你的生命有危险一样，你也不会相信浓烟与大火给予你任何理由具有这种信念。既然你不相信你犯了任何错误，你就不可能纠正错误。既然你不相信你具有那些认知理由，你就不能出于那些理由而选择相信你的生命有危险。

我们反过来假定，你确实相信，浓烟与大火给予决定性的理由让你相信生命有危险，那你也大可能由此选择相信你有生命危险。在绝大部分情形中，在相信我们具有决定性的认知理由要持有某个信念的过程中，我们也就有了该信念。而一旦我们已经有了某个信念，我们就无法再选择具有它。

例外也可能存在。我们接下来假定，尽管你相信浓烟大火给予决定性的理

由让你相信生命有危险，但你还是没有该信念。我们也许可以设想，你能由此选择让你自己出于这些理由而具有该信念。但是你对这些认知理由的回应仍只是部分程度上自愿的。当你看到浓烟大火正沿着旅馆楼梯上升时，你并没有选择相信：这些理由给予了你决定性的理由去相信你有生命危险。

我们为什么对于大部分认知理由的回应不可能是自愿的，对此还有其他的理由。我们要拥有对周边世界的知识，我们的信念就必须是可靠地引起的，由我们的视觉或其他感知经验或者是通过认识到其他事实（这些事实给予我们持有这些信念的认知理由）。如果我们能够自由地选择我们的一切信念，这种因果性就不可能是可靠的。我们要具有必然真理的知识，诸如逻辑真理或数学真理，我们就必须以某种理性但非自愿的方式对某些认知理由做出回应：比如承认或认识到从什么可以推出什么，或什么必定是真的。

50 这同样适用于我们的欲望和偏好。我们既不能选择我们有理由想要的事物，也不能选择这样的理由有多强，因此很少能选择我们想要的或偏爱的事物。我们能够选择的仅仅是把何种欲望作为目标，并试图去实现它们。我们对于这些理由的回应也许能比现在变得更自愿一点。这种自愿性的增强在某种程度上是种改善，因为我们由此就能够使自己变成我们有理由想要成为的那类人，从而我们的欲望、态度和情绪就能更容易得到转变。例如，我们也许能够确保，我们永远不会失去我们青年时的理想。但是这样的能力也有危险性，就像我们最近发现的以机械方式高速地移动我们的身体一样。如果自我的改变是变坏，那么我们刻意选择的新欲望也许会导致我们无法纠正的错误。

5 状态给予的（state-given）理由

我说过，我们有某种欲望的理由是由“欲望对象或者我们想要的事件”的相关事实所提供的。我把此种理由称为对象给予的。许多人认为我们还有状态给予的理由持有某种欲望。这种理由也是由某种事实给予的，只是不是由欲望对象的事实，而是我们持有某种欲望之状态的事实。不管是作为目的的还是作为手段，若我们持有某种欲望在某种意义上是好的，那我们就有这样的理由。

基于这种观点，我们拥有某种欲望的理由至少有四类，可以描述如下：

	目的式的或内在的	工具性的
对象给予的	我们想要的事件自身就是好的或值得获得的	这个事件具有好的效果
状态给予的	“我们想要某事”这个欲望本身是好的	“我们想要某事”这个欲望具有好的效果

我们可能同时有这四类理由去拥有同一欲望。例如，若你身处痛苦，我也许有四类理由想要结束你的痛苦。我想要的东西自身就是好的，它还有使得“你能重新享受生活”这种好的效果。“结束你痛苦”这个欲望可能本身就是好的，这个欲望可能还有好的效果，诸如你能为我的同情而感到安慰。

这对于我们持有信念的理由也同样适用。由于我们的认知理由与我们信念内容的真值相关，这些理由也可称为对象给予的。许多人也认为，我们还能具有状态给予的理由来持有某些信念。这些理由产生于这样的事实：该事实使我们对某种信念的拥有在某种意义上是好的。例如，常常有人主张我们有理由相信：上帝是存在的和我们会有来生。这些理由不是认知的或真值相关的，而是善性相关或基于价值的。具有信念的此类所谓理由有时候被称为实践理由或实用理由。

51

即使我们有这种状态给予的理由，我也认为这些理由根本不重要。我们若处于某种状态会更好，就有理由想要处于这种状态。如果我们能使自己处于这种状态，我们就有理由这样做。我们根本不值得去说，除了有理由想要和使自己处于这种状态外，我们还有理由处于这种状态。例如，假定若我苗条一点、有辆自行车、会跳舞和有些朋友，这会让我更健康 and 更幸福。这些事实就会给我理由去减肥、买自行车、学习跳舞，去交些朋友。根本不值得说，这些事实除了给我理由这样做之外，它还给我理由苗条、有辆自行车、会跳舞和有些朋友。这些理由毫无影响。

接下来假定：尽管我们若处于某种状态会更好，但我们不可能使自己处于此种状态。我们还是希望自己能处于这种更好的状态。例如，我可能有理由希望自己更高十英寸、再年轻20岁，速度比印度豹还快。我们没必要说我们还有理由处于高10英寸、能年轻20岁和速度能比印度豹还快的那种状态。此类说法显然是错误的。理由至少是有些人有能力做出回应的，但没有人能对年轻20岁这种理由做出回应。

这也同样适用于我们的信念和欲望。如果我们有某种信念或欲望对我们会更好，那么我们就有对象给予的理由想要这种信念或欲望，并尽可能使自己拥

有它们。根本不值得再说我们还有状态给予的理由来拥有这种信念或欲望。正如我在附录A中所论证的，还有其他的理由拒绝此类说法。

6 享乐式理由

52 “我们想要某种可能事件”这种对象给予的理由全由有关此事件的事实所提供。若这种理由是由“使得这种可能事件作为目的是好的，或者为其自身缘故就值得获得”的事实所提供的，它们就是目的式理由。若这种理由是由“某些事件具有好的效果，也就是作为某个好目的的手段”的事实所提供的，它们就是工具性理由。

如果目的式理由是由某些可能事件的内在属性或特征，或是由这个事件本身所包含的内容所提供的，那么它们就是内在的。例如，我们也许有理由想要使某人觉得不那么孤独，想从山顶去看优美的风景，或者想要理解生命或宇宙的起源。我们还有外在的目的式理由想要某个可能的事件，这种理由是由有关此事件与其他事件之关系的事实所提供的。但此种理由没有必要单独考虑，因为这样的事件会通过使某个更长的事件系列得到内在的改善，从而它作为其中的一分子而成为外在地好的。

何种事实给予我们内在理由？不同的客观理论对此有着某种程度上的分歧。这些理论既可能会诉诸有关福祉或者何种生活是最值得过的不同观点，还可能在我们有理由关注和促进谁的福祉上具有分歧。例如，按照理性利己主义来说，我们每个人有理由关注和促进的只是我们自己的福祉。按照理性的不偏不倚主义来说，我们总是最有理由平等地关注每个人的福祉。我认为，我们应当拒绝这两种理论。我们也不应该假定，只有“有关自己或他人的福祉”的事实才能提供对象给予的理由。还有其他的事物是值得拥有的。在数量繁多的对象给予的理由中，作为我们的例子而言，在此只考虑有关我们的享乐式福祉的事实所提供的理由就足够了。我认为，这些享乐式理由受到了广泛的误解。

53 在想要某事物之时，我们通常是对该事物的特征做出回应，这些特征给予我们想要该事物的理由。但是，我们具有某些类似于欲望的状态，它们并不以前面的方式来回应理由。饥、渴和性的本能状态是其中三例。还有一组重要的心理状态，尽管通常被认定是欲望，但最好还是被看作成一个独立的范畴。这些心理状态就是对于当下实际感知的那种享乐式喜好与厌恶，喜好使得我们具有这种感知是快乐的，厌恶就使得具有这种感知是痛苦的或以其他的方式是令

人不愉快的，感知是否令人快乐在部分程度上就在于这种好恶。

人们有时候主张，这些感知本身就有好坏之分，意思是其内在的定性特征或对它们的感觉给了我们理由喜欢或厌恶它们。但我认为，我们并不具有此种理由。这些喜好或厌恶也无所谓合理或不合理。这点在下面的情形下最为明显：对于同一种感知，比如我们吃牛奶巧克力、进行艰苦锻炼和冬泳所得到的那种感知，有些人喜欢，有些人则憎恨。有些好恶特别怪异。许多人讨厌刮粉笔的声音。我就讨厌摸天鹅绒的感觉，不喜欢家蝇的嗡嗡声，大多数顶灯的那种欲坠而让人麻木的效果。这些厌恶之怪异性并不减损我的理性。无论我们对某种感知是喜欢、厌恶还是漠不关心，我们既不是在回应也不是在不回应任何理由。

我认为，这也同样适用于许多审美体验。人们有时候声称，我们有理由欣赏伟大的艺术作品，为之感动或震撼。我认为，这在很多情形下都是错误的。我们有理由想要去欣赏这些艺术品，为之所感动或震撼。但是这些不是欣赏这些作品或被其感动或震撼的理由。我们的确有理由赞美某些小说、戏剧或诗歌，因为它们表达的观念极为重要。但诗歌在翻译中消失了，即使这个翻译表达了相同的观念。我们从来就没有理由欣赏伟大的音乐或者被其打动或震撼。如果我们问某些乐章如此奇妙的原因，答案可能是“远系调的三种调音”。这个答案描述了我们回应此音乐的原因，但不是理由。远系调的调音就像香草、调味品或其他使得菜好吃的各种成分一样。若某人不欣赏伟大的音乐或不为其所打动，此人的理性并不减少分毫，他并没有对理由不做回应。在把音乐与食物做此对比时，我并没有贬低音乐，也没有把它排在小说、戏剧和诗歌之下。尼采的说法貌似有理（尽管是错的）：没有音乐，生命将是个错误。但是，一般的审美理论大多在音乐方面栽了跟头。

54

这些说法是有争议的，因此我们可以回过来看看人们喜欢或厌恶的非美学感知。尽管这些感知本身没有好坏，但它们是那些有着好坏之分且更复杂的精神状态的组成部分。当我们疼痛时，坏的事物不是我们的感知，而是我们具有一种厌恶感觉的那种意识状态。如果我们不厌恶这种感知，那么我们的这种意识状态就不是坏的。这些感知给我们的感觉如何，在部分程度上取决于我们是否厌恶它们。有些人主张，若我们厌恶某种感知，这就会在某种方式上影响到这种感知的性质，那么这种感知本身就是坏的。基于这种观点，如下说法依然正确：如果我们不厌恶这种感知，那么无论是它们还是我们的意识状态，都不是坏的，我们也没有对这些理由不做回应。

当我们具有极喜欢或厌恶的感知时，大部分人会强烈地想要具有或避免具有这种意识状态。我们可以把有关这种意识状态的欲望称为元享乐式的（meta-hedonic）。许多人无法在享乐式好恶与这种元享乐式欲望之间做出区分。但这些心理状态有几个方面的不同。我们厌恶的是某种感知，而我们想要的是不具有为我们所厌恶的某种感知。这种欲望的满足要么是不再具有这种感知，要么是继续具有这种感知但不再厌恶。任何这样的主张都不适用于厌恶——与欲望不同，厌恶根本不存在是否被满足的问题。

另一种差别涉及时间。假定火正逼近我们的手，威胁着我们的是很快就会有巨大的痛苦。大多数人都极想避免这种临近的痛苦。但我们现在既不可能厌恶这种临近的痛苦，也不可能喜欢某种未来的快乐。与我们元享乐式欲望不同，我们的享乐式喜欢与厌恶既不可能指向未来，也不可能指向那种仅为可能的东西。这就是我为什么不把这些心理状态称为欲望的另一个理由。

55 如果我们要称这些状态为欲望，我们应该记住，由于这些状态与我们的欲望之间的这种差别，对这些状态为真的主张可能并不适用于其他的欲望。这些状态和我们的元享乐式欲望之间还有着其他重要且为人所忽视的差别。

首先，许多人相信我们的欲望能够创造或赋予价值或负价值。例如，科丝嘉写道，有些事物可以是“作为目的客观地好的，因为它是为了自身的缘故而被欲望的”。基于这种观点，我们通过重视事物而创造价值，事物因为对我们重要所以才重要。快乐与痛苦的例子似乎支持这种观点。正如我说过的，我们的享乐式好恶确实使得我们的有些意识状态有好坏之分。如果我们不能把这些好恶区别于我们的元享乐式欲望，我们就会相信这些欲望使得其对象出现好坏之分。这看起来就支持这种普遍的观点：我们的欲望创造价值。

科丝嘉的话就是一例。为了阐明其主张——某事物之所以能成为好的“是因为它由于其自身之故而可欲”，科丝嘉写道：“巧克力由它影响我们的方式而获得其价值。我们经由喜欢某物而赋予其价值”。我认为，这种例子并不能通过使得我们想要或避免的东西成为好的或坏的，从而表明我们的欲望能够创造或赋予价值。我们未来的快乐或痛苦或好或坏，这并不是由我们当下想有这些快乐或避免这些痛苦的欲望所造成的。当我们由于具有极为厌恶的感知而极度痛苦时，使得这种意识状态坏的是我们强烈的厌恶，而不是我们当下的欲望——不想要为我们所厌恶的那种感知。由于我们的元享乐式欲望并不使得其对象或好或坏，我认为，快乐与痛苦的例子并非决定性地甚或强有力地支持这种观点：我们的其他欲望具有这种创造价值的能力。尽管具有我们喜欢的那种感知

是好的，但没有任何事物仅仅因为我们想要它，从而它就是好的。

这两类心理状态之间还有一个重大的差别。与我们的享乐式好恶不同，我们的元享乐式欲望是对于理由的回应，因为我们有强有力的理由支持或反对具有这种欲望。这种差别足以表明我们应该区分这两类心理状态。当我们由于具有我们极喜欢的某种感知而体验强烈的快乐时，我们对于这种感知的喜好（to be liking）没有任何理由。如果我们并不喜欢这种感知，我们不会因此是不理性的或犯了错。但是我们有很强的理由想要具有且持续具有我们强烈喜欢的感知。我们甚至具有更强的理由不想处于痛苦之中，即不具有那种我们极不喜欢的那种感知，尽管这种不喜欢是没有任何理由的。

7 不合理的偏好

我已经说过，当我们想要的事件之特征能够给予我们想要此事件的理由时，我们的欲望就是合理的。当我们想要的某个事件是我们有理由不想要的，或者没有任何理由或只是很弱的理由想要的，我们的欲望就不是合理的，用老话说就是与理性相悖的。当我们的欲望清楚且强烈地与理性相悖时，这种欲望就是不合理的。其他此类欲望不过是不完全合理的。这里并不是泾渭分明的情形，因为不理性是个程度问题。

例如，假定我们必须在随后要遭受的两种折磨中做选择。如果一个折磨比另一个要痛苦得多，这个事实给了我很强的理由偏好较轻的那个。如果我们没有任何其他相关的理由，那么有意去偏好更痛苦的折磨就是有悖于理性的，从而是不理性的。

大部分这类偏好都涉及对时间的态度。首先考虑一个虚构的人，他有一种我们称之为“对未来周二漠不关心”的态度。这个人关心自己未来的苦乐，但若苦乐出现在周二，则完全例外。这个奇怪的态度并不依赖于无知或错误的信念。这个人知道，周二的痛苦与其他时候的痛苦是一样的，也同样是他的痛苦，且周二跟一周中其他日子也是一样的。即使如此，如果有选择，这个人还是会宁愿在任何周二受折磨，而不愿在其他日子受点小苦。折磨比小毛病痛苦得多，这点是不偏好折磨的强有力理由，但折磨发生在未来的周二，这根本不是偏好它的理由。因此，这个人的偏好与理性极为相悖，从而是不理性的。

接下来考虑某个人具有对来年的偏向。这个虚拟的人对于来年里的福祉一视同仁，但对于之后一年的福祉看得只有来年的一半重。他不是偏好来年第11

57 个月的5小时痛苦，而是偏好在12个月后的9小时痛苦。此类偏好也是不合理的。如果我们具有的痛苦不是一年之内而是之外，这根本不是把这个痛苦看得只有一半重的理由。

没有人对时间具有这种态度，但是，许多人都有我所谓对眼前的偏向（bias towards the near）。不像这两种虚构的态度，这种偏向并不会得出完全任意的区分。但假定：因为你有这种偏向，你想要你的折磨稍稍延迟，即使你知道这种延迟会使得你的折磨要糟糕得多。你偏好的不是今天一分钟的轻微折磨，而是明天一个小时的创痛。这种偏好依然是不合理的，尽管程度上要弱。许多人都基于不那么极端的此类偏好行事，由此使得其生活更差。

这些主张可能看起来太过明显而不值一提。有谁可能会否认创痛的性质会给予我们理由想要避免处于创痛中呢？又有谁会否认幸福的性质给予我们理由想要幸福呢？

然而，一些伟大的哲学家否认这样的主张，且许多晚近的合理性论说也否认这样的主张。正如我们马上会看到的，那些接受理由的主观理论的人必须否定这种主张。

第三章 主观理论

8 理由的主观理论

主观理论诉诸的事实有关我们当下的欲望、目标和选择。基于最简单的主观理论，我们可称之为 58

基于欲望的理论：我们有理由去做任何能满足我们当下欲望的事情。

然而，主观理论要具可行性，它们必须允许有些欲望无法给我们这样的理由。回到开头的那种情形：你想要逃离一条发怒的毒蛇，因为你误以为这样做会救你的命。如果你有理由满足你一切当下的欲望，那么跑开的欲望会给你行动的理由。但你没有理由跑开，因为静止不动才是活命的唯一方式。

有两种方法来解释：跑开的欲望为什么不能提供行动理由。主观主义者会主张

(A) 只有依赖于真信念的欲望才能提供理由。

(A) 蕴含着，你没有任何理由要跑开，因为你的欲望依赖于“这个行动会救你命”这个错误的信念。我们接下来回顾，当我们想要某个事件是作为目的，或者为其自身的缘故时，这种欲望是目的式的；当我们想要某个事件是作为实现目的的手段时，这种欲望是工具性的。我们的目标常常是我们决定要满足的目的式欲望。你想要跑开仅仅是作为救命的手段。这样，主观主义者反过来可以主张 59

(B) 只有目的式欲望或目标才能提供理由。

(B) 蕴含着，你没有任何理由要跑开，因为这个行动无助于你满足或实现任何这样的欲望或目标。

(A) 可能看起来比(B)更可行。若工具性欲望依赖于错误的信念，这看来使得这些欲望在某些方式是错误的，这可能成为这些欲望为什么不能提供理由的原因。这样的欲望若没有依赖于假(false)信念，则绝对不会是错误的。

然而，主观主义者可以用另一种不同的方式来捍卫(B)。假定我既想要吃某棵树上剩下的两个苹果，还想爬梯子以够到位置较高的那个苹果。假定树的主人只允许我吃一个苹果，但可以随便选一个。如果工具性欲望能给我们理由，那么我更有理由选择位置较高的那个苹果。如果我选择较低位置的那个苹果，那么我只能满足“我吃这个苹果”的欲望。如果我选择较高位置的那个苹果，我不仅能满足我吃另一个苹果的欲望，还能满足“我想爬梯子以够到那个苹果”这个欲望。但这个推理显然是错误的。因为我想爬梯子不是为了其自身的缘故，而只是为了够到那个苹果的手段，我没有进一步的独立理由来满足这个欲望。我要爬梯的理由，完全源自且丝毫无补于如下理由——满足“吃位置较高的苹果”的欲望。

正如这个例子显示的，工具性欲望并不提供理由。基于最简单且可行的主观理论，我们可称之为

目的式欲望理论：我们最有理由去做任何能够满足我们当下目的式欲望或目标的事情。

这个理论正确地蕴含着，你没有任何理由要逃离发怒的蛇。你的目标是救自己的命，逃跑无法实现这个目标。这里根本没有必要诉诸这个事实：你要跑开的欲望依赖于错误的信念。

60 然而，在有些情形下，我们的目的式欲望或目标依赖于假信念。例如，我可能想伤害你，原因是我错误地认为你罪有应得，或者是因为我误以为你伤害过我，而想报复你。主观主义者应当否认这种欲望给予我理由。主观主义者在考虑此类情形时，其中许多人主张，只有没有错误（意思是不依赖于假信念）的目的式欲望或目标才提供理由。按照我们所称的

无错的欲望理论：我们最有理由去做的事情要最有利于满足或达成我们当下无错的目的式欲望或目标。

显然还有其他方式可以改善这种理论。如果依赖于假信念的欲望无法提供理由，那么依赖于无知的欲望也无法提供理由。这个区分并不深刻。在我想伤害你的那个虚构情形中，有两种方式让我的欲望可能基于不当的根据。我可能错误地相信你有意伤害我，或者尽管“你伤害过我”这一看法是对的，但我不知道，你这样做是为了让我不受更严重的伤害。此情形的这两个版本并无实质差别。如果只要我伤害你的欲望依赖于假信念，则此欲望不能提供理由，那么欲望若依赖于无知时，同样不能提供任何理由。

如果基于无知的欲望不能提供任何理由，那么我们能言之成理地更进一步。我们若知道更多就不会具有现在的目的式欲望，则没有理由满足这些欲望，主观主义者据此能够主张：我们若知道更多就会具有某些目的式欲望，从而确实就有理由满足这些欲望。同前面一样，这个区分也不深刻。如果我得知你是出于好意而伤害我，那我可能就不仅不再对你恶意相向，而且会产生善意。主观主义者主张，如果此点为真，那么我现在有理由对你好。

如果我们诉诸知道更多而想要的东西，那么接下来就可以把这个观念推到极致，据

知情欲望理论：我们最有理由去做的任何事情，会最好地满足我们若知道一切相关事实后会具有的那种目的式欲望或目标。

61

有些作者主张，如果我们对某个事实的了解会影响我们的欲望，那么此事实就算作是相关的。但这个标准太过宽泛。正如吉巴德（Gibbard）所言，我们若知道且能生动地想象我们饭友的内脏里所发生的一切，就可能失去进食的欲望。我们若知道人对人何其不人道的某些事实，就可能非常压抑，不再想活下去。知情欲望理论由此就具有如下讲不通的蕴含：我们即使确实想吃且想活，也根本没有理由满足这些欲望。为了避免这样的含义，有些主观主义者主张，要把某个事实算作相关的，仅仅知道该事实会影响我们的欲望还不够。按照此类观点，若我们是在几个可能的行动之间做选择，相关的事实只是这些行动及其可能后果。

知情欲望理论还需要做一修正。情况有时候的确是这样：如果我们有充分

的信息，那么我们的欲望以及我们有理由做的事情都会发生变化，从而在某种意义上会改变我们的处境。如果主观主义者主张，我们的理由不是由我们的实际欲望提供的，而是由我们想象的知情欲望所提供的，那么这些人在此类情形中就会得出说不通的结论。例如，假定我们想要知道某些重要的事实。如果我们知道这些事实，我们就会失去这个欲望。但这一点并不应被看作蕴含了我们没有理由按这个欲望行事：努力了解这些事实。有些主观主义者因此主张，我们若是完全知情的，会设想自己在不知情的实际情况下会想要的欲望，这些欲望才是我们应该满足的。

还有些主观主义者，他们不是诉诸什么会最好地满足或实现我们的欲望或目标，而是诉诸在仔细考虑这些事实后我们会做的选择或决策。对于我们如何合理地做出这样的决策，这些人也提出了主张。按照我们可称之为的：

慎思理论：我们最有理由做的是在充分知情且理性的慎思后，我们会选择去做的任何事情。

62

这种形式的主观主义很容易混同为客观主义，因为其陈述方式可能给人以这些理论彼此相似的假象。主观主义和客观主义都可以主张：

(C) 我们最有理由或有决定性理由去做之事，等同于我们若是充分知情且理性的就会选择去做之事。

但这个主张是有歧义的。主观主义者和客观主义者都主张，我们若做出重大决策，则应当以某种方式慎思：我们应当充分想象各种不同的可能行动具有何种重要后果、避免一厢情愿的想法、正确地评估概率以及尊重其他的某些程序规则等。我们若这样去慎思，就是程序上理性的。

如果我们也是实质上理性的，那么对于我们具有的欲望和目标，还有我们可以做出的选择，客观主义者就会提出进一步的主张。这些主张是实质性的，这是在如下意义上说的：它们不是有关我们如何做出选择，而是有关我们选择的是什么。客观主义者认为，存在着多种目的式的欲望和目标，它们是我们所有人都具有强有力且通常是决定性的、对象给予的理由去拥有的。要成为实质上完全理性的，我们必须拥有这些欲望和目标并尽可能努力实现它们，由此对这些理由做出回应。慎思主观主义者不会提出任何这样的主张，而是否认我们

具有这种对象给予的理由，仅仅诉诸有关程序理性的主张。

这两类人尽管都会接受（C），但解释（C）的方式是不同的。按照主观主义者而言，当下述为真时，

（D）我们如果是充分知情且程序上理性的，那么就会选择以某种方式行动，

这个事实使得下述为真

63

（E）我们具有决定性的理由以这种方式行动。

客观主义者反过来主张，当下述为真时

（E）我们具有决定性的理由以某种方式行动，

这个事实使得下述为真

（F）我们若是充分知情且既是程序上理性也是实质上理性的，就会选择以这种方式行动。

为了阐明这些主张，我们可假定，除非戒烟，我就会早死，失去本可有的多年幸福生活。按照一切可行的客观理论来说，这个事实给予了我决定性的理由想要戒烟且努力戒烟。如果我是充分知情且是实质上理性的，这就是我会选择要做的事情。客观主义者相信，我们应当理性地选择的事情取决于我们有理由或显见理由想要且去做的事情。

我们接下来假定，经过完全知情且程序上理性的慎思（或我们现在可称之为理想慎思）之后，我会选择戒烟。慎思的主观主义者届时会同意我有决定性的理由戒烟。然而，基于这种观点，推理是反过来的：这些主观主义者不是主张我们的选择要取决于我们的理由，而是主张我们的理由依赖于这种慎思之后我们会选择的事情。如果我有决定性的理由戒烟，这是因为我会选择这种行动方式。

正如这些例子所显示的，这两种理论是完全不同的。客观主义者诉诸规范性主张：在理性的慎思后，我们有理由选择或应当理性地选择什么；而主观主

义者诉诸心理学主张：在这种慎思之后，我们事实上会选择什么。

不同的主观理论有时候会在我们有理由做什么上产生分歧。我们在此可以忽略这种分歧，只考虑这些理论一致的情形。在这样的情形中，我们知道一切相关事实，那么能最好地满足我们当下的目的式欲望或目标的行动，也是我们在理想慎思后会选择的行动。我们由此能说，根据

64 理由的主观主义：某个可能行动

是我们最有理由去做的，或者在蕴含决定性理由的意义上应该或应当去做的。

当且是因为

这个行动能最好地满足我们当下充分知情的目的式欲望或目标，或者是我们在理想慎思后会选择去做的。

某些主观理论之间还有一个分歧，我们可以在此提出但随后予以搁置。假定，鉴于相关事实，所有主观理论都蕴含着：我具有决定性的理由戒烟。有些主观理论主张，这个理由是这个事实给予的：

(1) 这个行动能最好地满足我当下充分知情的目的式欲望或目标。

按照其他某种主观理论，这个理由是由这个事实给予的：

(2) 戒烟会延长我的寿命。

但是这些理论主张，(2) 给我这个理由仅仅是因为(1)也为真。我想戒烟的理由是由“这个行动会延长我寿命”这个事实给予的，但是这个事实之所以给予我理由，是因为我想实现这个目标。同样的主张也适用于这个事实：

(3) 经过理想的慎思后，我会选择戒烟。

按照慎思主观主义者来说，我们有决定性的理由去做的任何事情，是我们

在理想慎思之后会选择去做的事情。但是，我们无法言之有理地主张，我戒烟的理由是由如下事实给予的：经过这样的慎思，我会选择这么做。有些人因此主张，(2)是给予我理由的事实，但这个事实之所以给予我这个理由，仅仅是因为(3)也为真。

在评估主观理论时，我们只要考虑“这些理论蕴含着我们有理由去做什么”就够了，可以忽略它们“在何种事实给予我们理由”上的那些分歧。基于这些理论，理由是由某些有关我们的欲望、目标或选择的事实提供的——我在这么讲的同时将包含这样的意思：这些事实属于那些使“我们具有这些理由”为真的事实。

理由的主观主义现在非常流行。许多人理所当然地认为我们具有主体给予的理由。例如，科丝嘉写道，如果某个行动“是获得你想要之物的手段……没人怀疑这是一个理由”。威廉斯写道：“想做某事的欲望当然是去做它的理由”。在许多著作与论文中，主观主义甚至都不被称作是几种观点之中的最佳者，而仿佛是作为唯一可能的观点而提出的。因此，这种观点是否为真就极为重要。

9 人们为什么接受主观主义

我认为，我们应当拒绝一切主观理论，并且接受某种客观理论。我们的实践理由全部是由对象给予且基于价值的。

既然有这么多人认为，所有的实践理由都是基于欲望、目标或选择的，情况怎么可能如客观主义者所主张的那样：并不存在任何这样的理由呢？这些人怎么会错得如此离谱呢？

我们的理由基于我们的某些欲望、目标或选择的方式看起来有好几种，因而存在着几种可能的局部解释。第一，正如我说过的，我们想要的通常是值得去做的事情或获得的事物。在这种情形下，既然我们具有基于价值且是对象给予的理由来满足这些欲望，因此，两种理论至少是部分一致的。

第二，我们有这样的欲望，是因为我们认为我们有这样的理由。我们常常是由某个行动或结果（在蕴含理由的意义上）是好的或最好的这个信念所打动的。我们通常是受到这样的信念之驱使：某行为或结果在蕴含理由的意义上是好的或最好的。当我们的欲望取决于我们的这一信念——我们具有这样的理由——之时，我们可能没有在这些欲望与信念之间做出区分。

第三，有些人接受基于欲望的福祉理论。按照其中某些理论，我们某些当

66 下欲望的满足对我们而言本身就是好的。如果情况确实如此，我们就有了基于价值的理由来满足这些欲望。

第四，当我们描述我们的驱动性理由或我们为什么会做所做的事情时，我们能够正确地诉诸我们的欲望或目标。这可能会导致我们认定，我们的欲望或目标也能给予我们规范性理由。有些人并没有区分这两种理由。

第五，我们确实可以在一种肤浅的意义上主张，我们的欲望或目标给予我们规范性理由。例如，我确实可以主张我现在有理由离开会场，因为我想要赶上某趟火车，或者我的目标是赶上这趟火车，并且现在满足这个欲望或实现这个目标的唯一方法就是离开会场。但是这种基于欲望或目标的理由是派生的，因为这种理由之规范力量完全来自于这样的事实：给予我理由想要赶上这趟火车或具有这个目标的事实。如果我没有任何理由要赶上这趟火车或具有这个目标，那么我就没有任何理由要现在就离开。当我主张我们的欲望或目标不能提供任何理由时，我指的是一阶理由，即非派生的理由。

第六，当我们能够满足他人欲望，或帮助他们实现其目标时，这些事实可以给予我们非派生的理由去这样做。当他人具有他们根本没理由要具有的欲望或目标时，这些人可能没有任何理由去满足这些欲望或实现这些目标。但是我们可以有这样的理由。我们在帮助他人满足其欲望和目标时，我们尊重了他们的自主性并且避免了家长主义做法。他人的欲望、目标或选择在这方面就像选票一样，即使投票者根本没理由这样投票，也应给予完全同等的权重。许多人之所以接受基于欲望或选择的理论，是因为他们是民主论者、自由主义者或自由至上者，他们认为，其他人应该想要、选择或做什么，我们对此不应当指手划脚。例如，诺齐克就主张：基于价值的实质性理论“打开了从外部强加专制要求之门”。

67 第七，我们若有某个目标且认为，某个可能的行动是实现该目标唯一或最好的方法，则我们可能应当理性地这样去做。有些人认定，在这样的情形中，我们必定有理由去做我们理性上应当做的事情。但情况不是这样的。我们若错误地认为某个行动会实现我们的目标，则可能根本没理由那样去做。你尽管在理性上应当逃离发怒的蛇，但没任何理由这样去做。

第八，当人们主张我们有理由满足当下欲望时，他们想的通常是我们认为在将来会享受的那些活动或体验。这样的信念若为真，我们就有理由满足这些欲望。但是这些理由的提供者不是“我们会满足这些欲望”这个事实，而是“我们会享受那些未来的活动或体验”这个事实。如果我们不享受那些活动或

体验，那么我们就没理由满足这些欲望。当孩子想要的是某个他们随后得到却不喜欢的东西之时，他们的父母亲有时候就会说，“瞧，你并不真正地想要它”。这样的说法是错误的，因为孩子们确实想要这些东西，真实情况是他们的欲望并不给予他们理由。同样的主张也适用于这样的欲望，即，我们想要避免我们认为是痛苦或不愉快的东西。当人们主张我们的欲望给予我们理由时，他们心中真正想到的通常是这样的事实：“我们会喜欢什么，发现什么是痛苦的或不愉快的”。这种事实给予我们的理由是享乐式的而不是基于欲望的。

第九，有些人错误地认为享乐式理由是基于欲望的。当这些人思考痛苦或不舒服的感知时，他们没有区分“我们对当前感知的厌恶”和“我们对不要我们所厌恶感知的那种元享乐式欲望”。我早就主张，正是我们的厌恶使得我们的意识状态变成坏的，并给予我们理由想要结束这种痛苦或不舒服的状态。由于这些人没有区分开我们的厌恶和我们的元享乐式欲望，他们就会认为这种欲望给予我们这种理由。类似的主张也同样适用于快乐和其他某些或好或坏的意识状态。

第十，我们所持有的诸多行动理由，是我们若没有某种欲望就不会有的。但这些理由之提供，不是由于我们的行动会满足这些欲望这一事实，而是由于某些其他像这样的事实——该事实因果性地依赖于我们对这些欲望的拥有。例如，我们若有某个欲望就会使得满足这个欲望是令人快乐的。在许多情形中，这个事实仅仅是锦上添花，因为我们想要的那个东西自身就是值得实现的。当我们没有任何理由要有某种欲望时，这种情形就表现得最为清楚了。例如，我们玩某种游戏，比如说没有报偿且其结果完全取决于运气的游戏，此时我们根本就没有理由想赢。但是如果我们确实想赢，那么这就使得我们会享受赢，而这一附随的事实由此就会给予我们理由想满足这个欲望。

在描述这样的情形时，我们可以做出另一个区分。按照某些主观理论来说，某些事实给予我们理由的方式依赖于我们具有某个欲望。这种依赖是规范性的。例如，基于某些观点，我戒烟的理由是它会延长我的生命这个事实给出的，但是这个事实之所以给我理由，仅仅是因为我想实现这个目标。这个理由的规范力量被认为是源于我有这个欲望这一事实，因而该理由是基于欲望的。我刚刚所描述的基于价值的理由则完全不同。若某个欲望的满足会给予我们快乐，这个事实就给予我们基于价值的享乐式理由，从而去做满足这个欲望的事情。这种理由可能因果地依赖于我们有这样的欲望，因为这种行为可以给我们快乐仅仅是由于我们具有该欲望。但是这个理由在规范性上并不依赖于我们具有这个欲望。“某个行动给予我们快乐”这个事实就给予我们理由去那样做，无

论这个快乐是否因果地依赖于我们具有那个欲望。

我们还有很多其他的理由，它们在因果性上也依赖于我们有某个欲望。例如，欲望得不到满足会让人沮丧或分神。这样的事实给予我们理由去满足这些欲望。正如前面一样，这些可能只是锦上添花的理由，因为我们想要的东西通常是值得获得的。但这样的情形可以包含我们根本就没什么理由要有的欲望。例如，我们可能会分心于想了解或记住某件琐细之事，或者是具有某个带有强迫性的欲望。我有时会出于一种不带感情的古怪欲望去剪指甲，并为此而分心。摆脱这类欲望的最好办法就是满足它们。

69 接下来假定：我们必须在两个或更多的目标之间做选择，其中没有哪个目标比任何其他目标更值得实现。这样的一些例子是：在不同职业或研究项目之间做选择，或者是在为不同的援助机构或政治选举做义工之间做选择。如果其间的一个目标是最强烈地想要实现的，这个事实就可给予我们理由采纳这个目标。但是同样地，这些理由不是由“我们最强的欲望是实现这个目标”这个事实提供的，而是由依赖于“我们有这个欲望”的其他事实提供的。例如，若其中的一个目标是最有吸引力的，这就会给予我们理由认为实现这个目标有最大的回报。一想到实现这个目标，事先就可能给予我们快乐。而我们实现这一目标的强烈要求，会使得我们更容易做出为实现这个目标所需要的努力与牺牲。在最落魄、无精打采或绝望之时，我们可能会需要这样的欲望。与前面一样，给予我们理由采纳并努力实现该目标的，不是我们的欲望本身，而是其他的那些事实。

这也同样适用于我们的决策和目标。当我们决定努力满足某个欲望，由此满足它就是我们的一个目标时，这个决策可以给予我们进一步的理由来满足这个欲望，从而实现这个目标。但提供这个理由的事实并不仅仅是“我们已经做了这个决策，采纳了这个目标”。提供这个理由的事实还有：如果我们不按这个决策行动，我们实现这个目标的可能性更小，而荒废时间的可能性更大。然而，在某些情形下，两者都不对，因为我们最好的做法莫过于重新考虑决策。例如，如果我们半夜醒来，那么重新考虑采纳某个目标的决策可能就不像苦等着再入睡那么乏味。在这样的情形下，“我们已经采纳了某个目标”这个事实并不能给予我们任何理由保持这个目标并努力去实现它，因为这个事实并不给我们任何理由让我们不再改变心意来采纳某个其他的目标。

我们有许多理由去满足我们的欲望或实现我们的目标，这些理由的提供者不是“我们会满足这些欲望或实现这些目标”这个事实，而是其他依赖于欲

望或目标的事实。正如前面一样，当人们主张我们的欲望或目标给予我们理由时，他们心中真正想的通常是其他的事实。

欲望、目标或选择给予我们理由的方式看起来就是有这么多，因此，有许多人接受主观理论就一点也不奇怪了。这中间的许多人对于哪种事实给予我们理由，其实有着各种各样正确的或可行的信念，他们只是不明白，这些信念事实上并不支持任何主观理论。尽管这些人相信他们是主观主义者，但情况并不真是如此。当这些人做出主观主义主张时，他们错误地描述了自己的观点。

10 分析性主观主义

人们还有另外一种方式来接受理由的主观理论。我们可称一种规范性主张为实质性主张，若这种主张

(a) 陈述某物具有规范性属性

且

(b) 是我们对之能够具有分歧的，由此是有意义的 (significant)，或者能告诉我们一些我们并不知道的事情，由此是富含信息的。

这类主张的两个例子是“处于痛苦是坏的”以及“较少关注未来是不理性的”。

正如西季威克与康德所提醒的那样，当我们思考规范性问题时，我们容易受到某类主张的误导，这类主张看起来是实质性的但不过是隐藏的同义反复。用康德的话来说：

没有哪门科学像伦理学那样充斥着同义反复。

一种公开的同义反复两次使用同一语词，这种方式并不做出任何实质性的主张，只是告诉我们某物是其所是，或者是如果某物具有某种属性，那么此物具有此种属性。

两个这样的例子是这样的主张：

(1) 幸福就是幸福

71 以及

(2) 产生幸福的行动会产生幸福

有些公开的同义反复可以用来提出有意义的主张。两个这样的例子是：“生意就是生意”和“战争就是战争”。当人们做出这样的主张时，他们是意图提醒我们某物极为独特，不同于其他事物，必须就其自身来判断。这些人可能意在暗示：在生意和战争中，日常的道德标准不适用。这些暗示性的主张是实质性的。但是大部分公开的同义反复是微不足道的。“幸福就是幸福，欲望是欲望，信念是信念，希望是希望”，这类主张没什么价值。

隐藏的同义反复不会使用相同的语词，而是使用具有相同意义的不同语词或短语。一个例子是这样的主张：

(3) felicity (幸福) 是 happiness (幸福)

由于“felicity”就是“happiness”，(3)所指的与(1)是一样的。(3)并不是一个实质性的主张，尽管我们可能会用(3)来告诉某人“felicity”的意思。接下来我们考虑这个主张：

(4) 产生幸福的行动是带来幸福的。

既然“带来幸福”意指“产生幸福”，所以(4)是另一个隐藏的同义反复，这两种公开的形式就是：

(2) 产生幸福的行动会产生幸福

和

(5) 带来幸福的行动是带来幸福的。

正如前面一样，这些都不是实质性的主张。每个理解这些主张的人都会接受它们，因为它们显然为真。每个人都能一致地接受这些主张，不管他们还相信其他的什么事情。(4)在这些方式上不同于这个主张。

(6) 产生幸福的行动是善的 (good)。

既然“善的”并不意指“产生幸福”，(6)就是有意义的实质性主张，有时候会与许多人的信念相冲突。例如，许多人认为，能给性虐待狂带来幸福的残酷行动绝对不是善的。

我们现在回到理由的主观理论。有些人是在我们称为主观主义或内在的意义上使用“理性”、“应该”和“应当”等语词。我们可把这些人称为分析性主观主义者。例如，当某些人说：

(7) 我们最有理由以某种方式行动。

他们的意思是：

(8) 这个行动能最好地满足我们当下充分知情的目的式欲望。

我们可称“最有理由”这种短语的主观主义意义是欲望满足的意义。有些这样的人还主张：

(9) 我们最有理由去做能最好地满足我们当下充分知情的目的式欲望的事情。

既然这些人在欲望满足的意义上使用“最有理由”这个短语，(9)就不是一个实质性的主张，而是隐藏的同义反复，其公开形式之一就是这样的主张：

(10) 会最好地满足我们当下充分知情的目的式欲望的行动，就是会最好地满足我们当下充分知情的目的式欲望的行动。

不管人们相信什么，他们都可以相信这种微不足道的主张。这也同样适用

于有关“理性”、“应该”和“应当”的其他主观主义的或内在的涵义。尽管对我们有理由去做什么或我们应当去做什么，分析性主观主义者并不提出实质性的主张，但这些人确实做出了其他重要的主张，我将在第24和30章加以探讨。

73 理由的主观主义者要做出实质性的主张，就必须在我们第一节所讨论的那种不可定义的规范性意义上使用“理由”、“应该”和“应当”这些语词。我在这些章节中探讨的正是这些实质性且非分析性的主观主义理论。

这里我们只要考察不同的主观理论于其中取得一致的情形就够了。在此类情形中，我们会知道一切相关事实，而其中最有利于实现我们的当下目的式欲望或目标的行动，也是我们经过理想的慎思之后会选择去做的。当我们的慎思是充分知情且程序上理性的时候，它就是理想的。在讨论这些理论时，我提出的某些主张仅仅是有关基于欲望的理由，但其中大部分也适用于基于目标和选择的理由。

在做出这些主张之时，我是在广义上使用“欲望”这个词，它涵盖了任何这样的状态：受到驱动、想要某事发生且在某种程度上愿意尽可能使之发生。然而，我的主张并不适用于包含欲望的各种复合状态。例如，当我们爱某个人时，我们就受驱动以某种方式行动。我们极为关注这个人的福祉，想做最有益于他或她的事情。尽管我们爱某人在部分程度上在于具有这样的欲望，但我们有很强的理由去关注且努力促进我们所爱者的福祉。提供这样理由的不是爱某个人所包含的各种欲望，而是爱本身就是好的那种方式，以及有关我们与爱人之间关系的其他各种事实，比如：有关共有的过去、承诺或感恩的理由的事实，在罗曼蒂克或肉欲之爱中所包含的事实，或者是对父母、子女或其他近亲的爱。为了阐明这种区分，我可以假定我遇到几个陌生人，他们都需要我的帮助。如果我有强烈的欲望帮助其中的一个，也许是因为我喜欢她的长相，这至多能给我一个非常弱的理由，去帮助她而不是其他的陌生人。各种形式的爱完全不同于这样一种欲望。

11 痛苦论证

主观理论会有说不通的含义。假定：

74 在情形1中，我知道未来某个事件会让我痛苦一段时间。即使经过理想慎思后，我仍然不想避免这种痛苦。这种痛苦或不想避免这种

痛苦也不会阻止我实现其他的欲望或目标。

既然我没有任何“避免这种痛苦”的欲望或目标，所有的主观理论就都蕴含着：我根本没有理由想要避免这种痛苦，也根本没有理由尽力去避免该痛苦。

人们也许认为，这种情形是不可能的，因为除非我有强烈的欲望不想处于这种状态，否则我的心理状态不会是痛苦的。但这种反驳忽视了我们当前痛苦与未来痛苦在态度上的差别。尽管我知道，当我后来感到痛苦时，我有强烈的欲望不想处于这样的状态，我现在可能没有任何欲望去避免这种未来的痛苦。

接下来人们可能主张，我预测到未来我不想处于痛苦之中，这给了我基于欲望的理由现在想要避免这种未来的痛苦。但如果人们接受我们这里所考虑的主观理论，那么他们无法提出这种主张。这些人不会主张（由于他们的其他假设，他们也不能主张），有关未来欲望的事实会给予我们理由。

其他的某些理论会做出这种主张。基于价值的客观理由理论可能会与一种基于欲望的主观福祉理论相结合。按照这样的一种观点，即使我们现在不关注我们未来的福祉，我们还是有理并且应当去关注。这些理由是由“使得未来各种事件对我们或好或坏”的事实提供的，在此意义上，它们是基于价值的。但是如果像这种理论所主张的那样，我们的未来福祉在部分程度上在于满足我们将来的某些欲望，那么基于价值的这些理由将是按照使未来欲望得以满足的方式来行动的理由。我们同样可以主张：我们具有基于价值的理由去满足他人的欲望，因为这样可以促进他人的福祉。尽管这些理论主张我们有理由去满足这些欲望，但它们是基于价值的客观理由理论，与我们现在所考虑的基于欲望的主观理论极为不同。

我们还能设想一种时间上中立的基于欲望的理论。基于这种观点，我们在任何时候最有理由去做的行动，是任何“能最好地满足我们一生中所有欲望”的行动，不管这些行动对我们是好还是坏。按照一种类似的个人中立的理论，我们最有理由去做的行动，是任何能最好地满足所有人欲望的行动，不管这些行动对任何人是好还是坏。这些虚构的理论也完全不同于我们现在所考虑的主观理论。

75

按照这些理论来说，只有有关我们当下欲望、目标或选择的某些事实，才给予我们理由，或者说我们的理由依赖于这些事实。情形1假定，我已经仔细地考虑了有关我未来时段痛苦的一切相关事实。既然这种痛苦或不想避免这种痛苦不会影响我们当前任何欲望或目标的满足，那么一切主观理论都蕴含着：我根本

没有理由去避免这种痛苦。这对于我的行动也同样为真。即使我轻易就能避免这种痛苦（也许是把我的手挪开正靠近的火焰），我也根本没理由那样做。这处理由的提供者只能是某种相关的当下欲望，而我没有任何这样的欲望。

某些分析性主观主义者会接受这种结论。如果这些人主张我没有任何理由要避免这种痛苦，他们的主张就不是规范性的，而是隐藏的同义反复，它只是重复了我对这个虚构情形的描述。这些人的意思不过是：在理想的慎思后，我没有动机把手挪离正靠近的火焰。我们都同意，在这种无足轻重且有误导性的意义下，我没有任何理由那样去做。

我们正在讨论的是非分析性主观主义者的观点。这些人是在规范意义上使用“一个理由”这个术语，它也可以用“算作支持”这个术语来表达。这些主观主义者同意：我有理由想要且尽力避免这种未来的痛苦，这样主张是有意义的。但是这些人的理论意味着：既然我没有任何相关的当下欲望，我就没有任何这样的理由。没有任何事实算作支持我想要避免或尽力避免这种痛苦。这也同样适用于其他的情形。按照这些主观主义者来说，我们若没有任何相关的当下欲望，则没有任何理由想要避免或尽力避免未来时段的痛苦。

我们现在能论证：

76

我们全都有理由想要避免或尽力避免一切未来的痛苦。

主观主义蕴含着我们没有任何这样的理由。

因此，

主观主义是错误的。

我们可称此为痛苦论证。

有些主观主义者可能主张，我们可以忽略这个论证，因为我的例子纯属虚构。他们会说，现实中的人都想避免一切未来的痛苦。

这个回应是失败的。第一，我们在问的是：主观理论是否蕴含我们都有理由想要避免一切未来的痛苦。要主张我们都有这种理由，只主张每个人具有这种欲望是不够的。这些主观主义者还必须主张：当我们具有某个欲望时，这个事实给予我们理由具有这种欲望。正如我们马上会看到的，这个主张得不到辩护。

第二，现实中的人看来有可能并不想要避免一切未来的痛苦。痛苦若在遥

远的未来，许多人就极少关心。有些人即使相信罪人会在地狱受苦，也仅当病人膏肓、离地狱不远时才力免犯罪。并且有些人在极为消沉之时就不再关心自己未来的福祉。

第三，即使不存在任何这样的现实情形，规范理论在纯属虚构情形（只要这些情形涉及的事情是足够清楚的）中的蕴含也应当是可接受的。主观主义者提出了何种事实给予我们理由的主张。但这样的主张在现实世界中不可能为真，除非它们在如下的世界也为真——如此世界中的人与我们很相像，只是他们不想避免一切未来的痛苦，或者他们的欲望与我们的有所不同。我们因此就可考察虚构情形，从而公平地检验主观理论。

77

主观主义者可能会回应道，即使在这样的可能世界里，每个人也必定具有某种目的式欲望，因为没有这样的欲望，他们甚至不可能是理性行为者，即不可能是出于理由而行动的人。威廉斯表示，我们若要做这样的行为者，则必须具有一种“不会因错误而失败的欲望”，并且具有“适度的审慎”。但是这样的主张在此是不相关的。我们可以是出于理由而行动的行为者，且不用想避免一切未来的痛苦。

主观主义者接下来可能还会主张，如果某个理论在一切或大部分实际情形中具有可接受的含义，这个事实就可以给予我们充分的理由接受这种理论。即使在有些不同寻常或虚构情形中这种理论具有错误的含义，我们接受这种理论还是可证成的。

许多理论能可行地以这种方式得到捍卫。然而，这种捍卫要想成功，我们必须能主张并不存在这样的一种理论，它与之相竞争，且具有更可接受的含义。主观主义者不能做出这样的主张。当主观主义理论应用于实际的人时，这些理论常常具有可行的含义。但这是因为大部分人通常具有的欲望是他们有对象给予的理由要具有的，因为他们想要的东西在某种程度上是好的，是值得获得的。在很多此类情形中，主观理论与最好的客观理论具有相同的含义。在努力想判别何种理论最好时，我们必须考虑两种理论具有不同含义的情形。判别不同科学理论的方式也是出于同样的理由的。这些分歧在不同寻常的情形或虚构情形中表现得最为清楚。因此，主观主义者不能主张，我们可以忽略这样的情形或不应当那么看重它们。相反，这些情形恰恰是我们最有理由要去考虑的。我一直在论证，主观理论有关这些情形的主张远没有最好的客观理论那么说得通。如果只要两类理论出现分歧，客观理论就更可行，那么它就显然更好。

还有另一种可能的回应。慎思主观主义者诉求我们经过知情且理性的慎思

78

之后想要或选择的事情。这些人可能会主张：

(A) 我们所有人都有理由持有这样的欲望：该欲望是任何完全理性的人都会具有的。

(B) 任何充分理性的人都想要避免未来的痛苦。

因此

我们所有人都有理由想要避免一切未来的痛苦。

然而，正如我说过的，这样的主张是有歧义的。客观主义者能够接受 (B)，因为这些人提出了有关实质理性的主张。按照客观理论来说，我们所有人都有决定性的理由具有某些欲望，而要成为实质理性的，我们就必须具有这些欲望。我们想要的事物或不想要的事物，它们的内在特征给出了这些理由。我们具有这样一种决定性的、对象给予的理由想避免一切未来的痛苦。如果我们并不具有这种欲望，我们就不是彻底实质理性的，因为我们没有对这种理由做出回应。

然而，主观主义者不能做出这样的主张。基于主观理论，我们并没有任何对象给予的理由，甚至没有理由想要避免一切未来的痛苦。慎思主观主义者诉诸我们在慎思之后想要的事情，这仅仅是程序上理性的。基于这些理论，如果我们具有某些目的式欲望或目标，那么我们受到的理性要求可能是想要且去做能最好地满足我们的欲望和实现我们目标的事情。但是，也许除了几种欲望——我们若缺乏它们甚至就不可能作为行为者——之外，没有任何目的式的欲望或目标是理性要求我们具有的。无论我们关注什么或想要实现什么，我们都可以是程序上理性的。正如罗尔斯写道，一个主观主义者

知道人们是理性的，但我们并不知道他们会追求什么目标，仅仅是知道他们会明智地去追求。

79 因此，主观主义者无法主张，任何充分理性的人都想要避免一切未来的痛苦。

也许有人会反驳道，我在说这些话时低估了主观主义者通过诉诸有关程序理性的主张所能取得的成就。例如，史密斯主张：

(C) 理性要求我们：不具有做出某种任意区分的欲望或偏好。

史密斯写道，主观主义者通过诉诸这种“最小原则”，他们能够解释许多欲望或偏好的不合理，例如我虚构的那个人的偏好：他关注会发生在他身上的一切痛苦，但周二除外。史密斯主张，这个人的偏好是不合理的，因为他做了任意的区分。主观主义者还可能主张，不想避免一切未来的痛苦也是同样任意的。

然而，主观主义者不能做出这样的主张。只有当且因为我们所偏好的事物绝对不是更可取的时候，我们的偏好才做出了任意的区分。如果有关两个事物的任何事实都不能给予我们理由偏好哪一个，那么偏好其一就是任意的。我的虚构者会偏好具有两种类似的折磨之一，只要且因为这种折磨是在周二。要解释这个偏好为什么是任意的，我们必须主张：

(1) 如果某种折磨在周二，这个事实并不给予我们任何理由不那么看重它。

与我的虚构者不同，我们大部分人会总是偏好两种折磨之一，只要且因为这个折磨带来的痛苦较轻。要解释为什么这个偏好不是任意的，我们必须主张：

(2) 如果某个折磨的痛苦较轻，这个事实确实给予我们理由不那么重视它。

(1) 和 (2) 是有关对象给予理由的主张。既然主观主义者否认我们具有这种理由，这些人不可能诉诸这种主张，或者诉诸史密斯在 (C) 中所陈述的“最小原则”。

史密斯还主张：

(D) 理性能要求我们具有某个欲望，当且因为我们具有这种欲望将使得我们的欲望组合更为融贯和统一。

为了阐明这个要求，史密斯假定，对于我们所知道的身处绝境者，我们仅仅想要帮助其中的某些人。史密斯主张，如果我们想帮助所有这样的人，我们

的欲望就会更为融贯，会“更加说得通”。

(3) 只要某个人极为贫穷，这个事实就给予我们理由想要帮助他。

如果这个事实并不给予我们这种理由，且如果我们只想帮助其中的一些人，则我们的欲望将不那么融贯，或者不那么说得通。而(3)是另一个有关对象给予理由的主张，是主观主义者无法诉诸的。

我们接下来考虑史密斯的这个主张：理性能要求我们具有更统一的欲望组合。单纯的统一性并不是一个优点。如果我们是偏执狂，只关注一件事情，我们的欲望将更为统一。但是如果你关注真理、美与人类的未来，而我只关注集邮，那么史密斯的主张似乎蕴含着，你的欲望没有我的那么合理。史密斯可能回应道：若我具有几种融贯的欲望，则我的欲望组合的统一让人印象更深刻。但如果我还想收集火柴盒、图钉、票根和塑料杯，你不那么统一的欲望组合还是比我的要更合理。这种对融贯性的诉诸再次假定，我们具有对象给予的理由要具有我们的欲望。主观主义者否认我们具有这种理由。

还存在着其他的问题。假如我们现在不关注未来的某些痛苦，但我们如果不关注未来的任何痛苦，那么我们的欲望就会更融贯。出于这一切理由，主观主义者不能主张：如果我们是程序上理性的，我们就会想要避免一切未来的痛苦。

81 既然主观主义者无法捍卫这个主张，我早前的结论就是站得住脚的。主观主义者必须主张，在情形1中，我没有任何理由要避免我未来时段的痛苦。正如我说过的，我们能论证：

我们所有人都有理由想要避免且尽力避免一切未来的痛苦。

主观主义蕴含我们没有任何这样的理由。

因此

主观主义是错误的。

有些主观主义者现在可能会咬紧牙关，否认我们有这种理由。这些人可能会说，在情形1中，尽管靠近的火焰会使我们难以忍受的痛苦，但这个事实并不能算作支持我们想要且尽力把我们的手移开。但这令人难以置信。

我们接下来可以回顾主观主义为什么具有这样的含义。既然主观主义者否认我们具有对象给予的理由，他们就必须同意，基于他们的观点：

(E) 痛苦的性质并不给予我们任何理由想要避免处于痛苦之中。

我们能主张

痛苦的性质确实给予我们这样的理由。

因此，主观主义是错误的。

我相信，这些论证是决定性的。

主观主义者可能会抗议，我们在否认(E)时并没有用论证反对他们的观点，只不过是拒绝了他们的观点。情况若是这样，则我们的主张毋宁是每个人都应当拒绝这种观点，因为(E)是一种完全说不通的信念。主观主义者不是虚无主义者，后者否认我们具有任何理由。主观主义者认为我们具有行动的理由。如果我们能具有某种理由，那么如下事实是再清楚不过了：在蕴含理由的意义上，处在痛苦中是坏的。我们可能很难准确地记起，具有一种极为痛苦的感知是何种状态。有些极糟糕的感知消失了，但我们能足够好地记起这样的体验。按照主观主义者来说，我们的记忆不能给予我们任何理由想要避免再次身处此种强烈的痛苦之中。如果我们问“为何不能？”，我相信，主观主义者根本无法给出令人满意的答案。

82

第四章 进一步的论证

12 全部或全不 (all or none) 论证

83 我说过，我们有理由持有某种目的式欲望，例如有理由想要避免一切未来的痛苦。我们现在可以问，我们的目的式欲望，是否如主观主义者所主张的那样会给予我们理由。

假定，在

情形2中，我想要拥有某种未来时段的痛苦。我不是一个受虐狂，后者想要这种痛苦是作为得到性快感的手段。我也不是一个忏悔的罪人，要把这种痛苦作为我的罪有应得。我当下也没有任何欲望或目标，需要经历这种未来时段的痛苦才得以满足或实现。我想要这种痛苦是作为目的的，或者说为了这种痛苦自身的缘故。我也没有任何这样的当下欲望或目标，它们的实现要么被这种痛苦所阻止，或者是被我对这种痛苦的欲望所阻止。在理想的慎思后，我决定，我尽力让自己在未来具有这种痛苦。

主观理论在此蕴含着：我具有决定性的理由要满足我的欲望，且根据此决策而行动，使我自己处于痛苦之中。如果附近有火，且我没有任何其他的方式能满足我的这个欲望，那么我就具有决定性的理由把手伸到火里去。这令人难以置信。

84 主观主义者在回应这种反驳时说，我们无法融贯地设想情形2。有些作者主张，如果我们确实相信：随后处于痛苦中的是我们，并且也理解这种痛苦会

是什么样子的，那么我们想要自己后来处于这种状态是不可想象的。但这个主张是错误的。我们想要对我们明知是坏的东西，这是可能的。某人为了痛苦自身的缘故想要在某个未来时段具有痛苦，这个假定是说得通的。主观主义者也不能这样主张：我们若有这个欲望就会使得我们不可能是理性的行为者，即出于理由而行动的人。

某人可能为了痛苦自身的缘故而想要未来的痛苦——这种情形尽管是可设想的，却是难以想象的。这个事实似乎弱化了这个反驳对主观主义理论所具有的力度。

反过来才是对的，这个事实加强了反驳的力度。如果我们发现很难想象有人会有这种欲望，恰恰是因为我们接受了客观主义的主张。我们认为，痛苦的性质给了每个人以强有力的理由不想处在这种状态之中。而按照主观理论，我们没有任何这种对象给予的理由。如果的确如主观主义所说的那样，那么就不会难于想象某人会为了痛苦自身的缘故而想未来时段处于痛苦之中。我们最多能够主张：这种欲望是不同寻常的，如同某些人的古怪性趣味。我们难于设想这种情形，恰恰是因为痛苦的可怕给予每个人以清晰且强有力的理由不具有这种欲望。很难相信人们会如此不理性。

主观主义者为了回答这种反驳，可能会修正其观点。他们可能会主张：

(F) 某些欲望或目标要给予我们理由，我们就必须具有理由要具有这种欲望或目标。

主观主义者只要能诉诸(F)，那么就能主张：既然在情形2中我没有任何理由想要在未来某个阶段受苦，他们的理论就没有这样的蕴含，即，我有满足该欲望的某种理由。

为评价这种回应，我们能假定，在

情形3中，我想要避免未来时段的痛苦。

主观主义者能主张我有理由想要这种欲望吗？在我们的例子中，我们是在假设：我们知道一切相关事实，并且已经经历了理想的慎思过程。主观主义理论蕴含着，在这样的情形中

(G) 我们要有理由具有某个欲望或目标，我们就必须具有某个当下的欲望或目标给予我们这个理由。

我们有一种直截了当的方式来主张，我们具有某个基于欲望或目标的理由来想要避免某个未来时段的痛苦。主观主义理论蕴含着：

(H) 若某个可能的事件有我们想要的某些效果，或者能帮我们实现某个目标，这个事实就给予我们理由想要这个事件，以作为达到这些效果或实现这个目标的手段。

假定，如果我今天下午参加棋类比赛时头痛，那么头痛就会让我分心，从而得不到想要的胜利。主观主义理论蕴含着：我有理由想要避免头痛，以作为帮助我赢得这场比赛从而满足我欲望的手段。但我们可以假定：在情形3中，我没有任何这样的工具性理由想要避免那种未来时段的痛苦。既然这种时段可以是非常短暂，避免这种痛苦就不会具有我想要的任何其他效果，或者能帮助我们实现当下的目标或满足当下的欲望。基于这些假设，(H) 蕴含着我没有任何理由想要避免这种痛苦。

主观主义者也可以主张

(I) 若下面两者之一为真

(a) 我们对某个欲望或目标的拥有会有我们想要的效果；或

(b) 我们想要拥有这个欲望或目标。

86

这些事实就给予我们理由拥有这个欲望或目标，或至少给予我们理由尽力具有或保持这个欲望或目标。

但在情形3中，我可能没有任何这样的理由。首先假定我无法避免未来时段的痛苦。部分程度上是由于这个理由，我想要避免这种痛苦的欲望达不到我想要的效果；并且这个欲望会有我不想要的效果，因为它让我对未来要发生的事情充满焦虑。出于这些理由，我不想要这个欲望。基于这些假设，(I) 并不蕴含着我有任何理由去持有或保持这个欲望。

既然我当下的其他任何欲望或目标，都无法给予我基于欲望或目标的理由，让我想要避免这种痛苦，那么主观主义者现在可以宣称，这个欲望本身给

了我这种理由。主观主义者为了捍卫这种主张，可能会说

(J) 我们若有某个充分知情的当下欲望或目标，这个事实就给予我们理由想要这个欲望或目标。

如果 (J) 为真，所有这种欲望或目标将是合理地自我证成的 (self-justifying)。我想要避免这种痛苦的欲望给予我理由想要这个欲望。如果我想处于痛苦之中，这个事实会给予我理由处于这种痛苦之中。如果我想荒废生命，这个事实会给予我理由想要荒废生命。不管我们想要什么，我们对这种知情欲望的拥有就会给我们理由想要它们。既然这样的主张显然是错误的，主观主义者就必须拒绝 (J)。由于主观主义者不能诉诸 (J)，这些人必须同意，在情形3的这个版本中，“想要避免未来痛苦”的欲望不能给予我任何理由想要该欲望。既然我当下其他的任何欲望或目标都无法给予我任何理由想要这个欲望，这些人现在就必须承认：基于他们的观点，我没有任何理由想要避免这个痛苦。

我们接下来假定，在此情形的一个不同版本中，我能避免这个未来的痛苦。我有这个欲望，由此导致我去做避免该痛苦的事情，从而满足该欲望。人们可能会主张，这个事实给予我基于欲望或目标的理由要拥有这个欲望。主观主义者可能会说

(K) 我们若持有某个知情欲望，则导致我们去做满足这个欲望的事情，这个事实就会给予我们理由要具有这个欲望。

87

但若 (K) 为真，则所有此种可满足的欲望都会是合理地自我证成的。如果我们想要处于痛苦中会导致把手伸到火中，这个事实就会给予我们理由想要处于痛苦之中。如果我们想要荒废生命会导致我们荒废生命，这个事实就会给予我们理由荒废生命。既然这些主张显然是错误的，主观主义者就必须拒绝 (K)。这些人必须再次承认，基于他们的观点，我没有任何理由想要避免未来时段的痛苦。这样主观主义理论就蕴含：在情形3的两个版本中，我都没有任何理由要具有（想要避免未来时段痛苦的）这个欲望。

我们有很多此类的实际情形。当我们想要避免某个未来时段的痛苦，或想要较轻的痛苦时，情况通常是：即使是经过理想的慎思，这个未来的痛苦也不

会阻碍我们满足当下任何其他的欲望或目标；并且无法主张，当下的任何欲望或目标会给予我们基于欲望或目标的理由想要避免这种痛苦。因此，主观主义理论就蕴含着，我们常常没有任何理由想要避免未来时段的某个痛苦。

这也同样适用于许多其他类型的情形。当我们想要自己或他人具有某个未来时段的幸福，或者我们具有其他好的或合理的目标时，情况常常是，即使在理想的慎思后，实现这些目标也不会满足我们当下任何其他欲望或目标；并且无法主张，存在其他任何的欲望或目标会给予我们理由要具有那些目标。情况之所以往往是如此，是因为我们想要这些东西是为了其自身的缘故，不是作为实现其他欲望的手段。这样一来，主观理论蕴含着我们常常没有任何理由，想要自己或他人具有这种时段的幸福，或者具有其他好的或合理的目标。

我们现在回到这个主张：

(F) 某些欲望或目标要给予我们理由，我们就必须具有某种理由去持有该欲望或目标。

88 我们已经在情形3中看到，我没有任何基于欲望或目标的理由要具有“避免未来痛苦”的欲望。因此，如果主观主义者接受(F)就得主张：“想要避免这个痛苦”这一欲望不能给予我任何理由这样做。即使我只要把手挪离正靠近的火焰就能轻易满足这样的欲望，我也没有任何理由这么做。这个主张与所有主观理论相矛盾，且显然是错误的。因此，主观主义者不能诉诸(F)。

还有另一个理由使得主观主义者无法主张：欲望要给予我们理由，我们必须得有理由要具有这种欲望。正如我们已经看到的，基于这些人的观点，任何这样的理由都必须是由其他欲望所提供的。(F)蕴含着，如果这个欲望要给予我们理由，我们就必须有理由来持有这个欲望。基于主观理论，这个理由还必须由其他某个欲望提供，由此直至无穷。我们不可能持有任何像这样无起始的、基于欲望的理由与欲望之链。按照这种理论，任何这样的链条都必须开始于某个我们没有理由具有的欲望，或以其为根据。这样一来，如果主观主义者诉诸(F)，他们就不得不得出：没有任何欲望能给予我们理由，由此就否认了这种理论的主要主张。

既然主观主义者不能诉诸(F)，他们就必须承认，基于他们的理论

(L) 我们最有理由做的事情，就是能最好地满足或实现我们当

下充分知情的目的式欲望或目标的事情，不管我们想要什么，我们是否有任何理由想要这些欲望或目标。

这也同样适用于我们在理想慎思之后所做的选择。

我们现在可以回到情形2，其中我想要某个未来时段的痛苦，不是作为手段，而是作为目的，或说为了其自身的缘故。这个未来的痛苦或我对这个痛苦的欲望，既不会阻止也不会满足我当下的任何欲望或实现当下的任何目标。在理想的慎思后，我决定尽可能使自己具有这种痛苦。既然主观主义者必须接受（L），他们就必须承认：基于他们的观点，我最有理由使得自己为了痛苦自身的缘故而处于痛苦之中。这样做能最好地满足我当下充分知情的目的式欲望，且是在理想的慎思后，我会选择去做的。如果附近有火，且我没有任何其他的方式满足我的欲望，我就会有决定性的理由把手伸到火中去。这个令人太难以置信。鉴于我对这种情形的描述，我认为，没有任何事实可以为“我把手伸进火中”提供哪怕是很弱的支持。我有决定性的理由使自己不这样处于痛苦之中。

89

类似的情形可能还有许多。按照主观理论，如果我们具有这种知情的欲望要击打我们正号啕大哭的娃娃，或者砸烂运转不畅的机器，这些事实就会给予我们理由，去击打娃娃或者砸烂机器。如果我们最想要或最想选择的是挫败我们所有未来的欲望，这个事实就会给予我们理由，去挫败我们所有未来的欲望。如果我们最想要的或最想选择的，就是荒废我们的生命，实现坏的或无价值的目标，那么这些事实就会给予我们决定性的理由，去荒废我们的生命，努力实现这些坏的或无价值的目标。这些主张也令人难以置信。我认为，主观理论的这些含义给予了我们决定性的理由，要拒绝一切这样的理论。

主观主义者可能回应道，尽管这些欲望和选择不能给予任何行动理由，这并不表明没有任何欲望或选择能给予我们理由。这些人必定承认，在情形2中，我想处于痛苦之中的欲望不能给予我任何理由这样做。但是主观主义者主张，在情形3中，我不想处于痛苦之中的欲望确实给予了我理由。这些人会做出类似的主张：如果我们想要的是其他方式的痛苦，或荒废我们的生命，或者实现其他坏的或没价值的目标，那么我们没有理由要满足这样的欲望；但如果我们想要的是幸福，过一种充实且有价值的生活，或实现其他的好目标，那么我们确实有理由满足这样的欲望。

然而，主观主义者不能提出这样的主张。这些主张诉诸这些欲望或目标的对象在给出理由的特征上的差别。如果我们做出这样的主张，我们就转到了

一种客观理论，诉诸对象给予的理由。主观主义者不能用这样的方式区分欲望或目标能否给予我们理由。在正予以考虑的情形之中，我们知道一切相关的事实。在这样的情形下，我们能够论证：

如果我们具有基于欲望的行动理由，要紧的就全然在于，某个行动是否会满足我们经过理想的慎思而当下持有的目的式欲望。我们想要什么或努力想实现什么是不相关的。

因此

要么一切这样的欲望，要么没有任何欲望，能给予我们理由。

如果一切这样的欲望给予我们理由，我们的欲望就能给予我们决定性的理由，为了其自身的缘故，使我们处于痛苦之中，荒废我们的生命，努力实现无数其他坏的或没价值的目标。

我们不会具有这样的理由。

因此

没有一个欲望能够给予我们任何理由。我们没有任何基于欲望的理由要持有任何欲望，或做任何事情。

我们可以称此为全部或全不论证。同样的论证也适用于基于目标或选择的理由。

当我们想要避免痛苦或过得幸福，或具有其他好的或合理的目标时，我们确实有理由来努力满足这些欲望和实现这些目标。但提供这些理由的不是“这些行为会满足这些欲望和实现这些目标”这一事实，而是我们当作目标去想要或拥有的事物之特性，这些特性使得这些事件是好的或值得实现的。

对于这个结论，这里还有一个重叠的论证。按照客观主义者来说，若某个事件或行为具有的效果是我们有理由想要的，那么我们就具有工具性的理由想要此事发生或做此行为。正如这个主张所蕴含的，每个工具性理由都是从其他某个理由获得其规范效力的。后一理由可能本身又是工具性的，是从第三个理由获得其效力的。但是在任何这样的理由链条的开始，必定存在某个事实给予

我们理由，想要把某个可能的事件作为目的，或者说为了其自身的缘故而想要它。这种理由是由某种内在特征所提供的，正是这个特征使得这个事件在某种方式上是好的。而一切工具性理由的规范效力恰恰是派生自这种对象给予的、基于价值的理由。

主观主义者必须拒绝这些主张。按照这些人来说，工具性理由的效力不是源自某个目的式理由，而是源自某个目的式欲望或目标。我们可以基于欲望的理由来持有某些欲望，并且我们可以具有由基于欲望的工具性理由和欲望组成的长链。但是在任何这样的链条的开端，正如我们看到的，必定总是存在某个欲望或目标是我们没有任何理由要具有的。正如我们的例子帮助我们看到的，我们不能言之成理地主张，这种欲望或目标给予我们理由。我没有任何理由把手伸到火中去，我们没有任何理由击打我们正号啕大哭的娃娃，荒废我们的生命，努力实现无穷的其他坏目标和无价值目标。由此看来，主观理论建基于浮沙之上。既然主体给予的理由之效力必须源自某个欲望或目标，而这样的欲望或目标又是我们没有任何理由要持有的，因此，我们就不能言之成理地主张：这样的欲望或目标能给予我们任何理由，我们具有主体给予的任何理由。我们不可能具有任何这样的理由，据之持有任何欲望或目标，或做任何行动。

13 不融贯性论证

主观主义者可能会再次抗议：我的论证诉诸的是纯粹虚构的情形。他们可能会主张，若应用于实际情形，主观理论的含义是可接受的。

然而，正如我说过的，好的理由理论必须能成功地应用于虚构情形。并且我也不只是诉诸此类情形。我也论证过，在许多实际情形中，主观理论蕴含着：我没有任何理由想要自己或他人：避免未来时段的痛苦、具有未来的幸福时光和许多其他的好目标。尽管主观理论通常具有的含义是可接受的，但这个事实并不能支持这些理论，因为只有当这些理论与最好的客观理论重合时才具有此类含义。

92

为了阐明第三点，让我们来比较两类认知理论。按照

基于理由的理论，我们应当相信，已知事实给予我们决定性理由要相信的事情。

按照一种不可行的虚构理论，我们可以称之为

基于信念的理论，我们应当相信，我们在权衡事实后实际上相信的任何事情。

当应用于现实中的人时，这种基于信念的理论具有的含义通常是可接受的。既然我们大部分人常常相信的事情，是我们所权衡的事实通常给予我们决定性理由要相信的，这种基于信念的理论就通常蕴含着：我们应当相信的事情，是我们有这种决定性理由要相信的。但那不是这种理论的主张。在其关于“我们应当相信什么”的主张中，这种理论蕴含着：我们没有任何理由拥有我们的信念。当这种基于信念的理论具有的含义是可接受的时候，这是因为现实生活中的大多数人认为，他们确实具有这种理由，且具有的信念也通常对这种理由做了回应。由此看来，我们应该拒绝这种理论。

这也同样适用于“有关我们应当做什么”的各种理论。按照我们这里可称为的

基于理由的客观理论，我们应当努力实现“已知事实给予我们决定性理由要具有”的目标。

按照

基于目标的主观理论，我们应当努力实现“在权衡事实后我们实际上会有”的目标。

当这些主观理论应用于现实生活中的人时，它们通常具有可接受的含义。既然我们大部分人持有的目标是“我们要权衡的事实会给予我们决定性理由”要持有的，主观理论通常蕴含我们应当努力实现这些目标。但这不是这些理论的主张。这种理论有关我们应当做什么事情的主张中，它蕴含的是我们没有任何理由持有我们的目标。当这种基于目标的主观理论具有可接受的含义时，这是因为大部分现实中的人接受他们确实具有某种理由，并且具有的目标通常对这种理由做出了回应。因此，我们应该拒绝这种理论。我们可能说，这些理论之所以看起来是讲得通的，仅仅是因为大部分人并不相信这些理论的主张。

许多主观主义者并不完全相信自己理论的主张。我们一直在讨论的都是知道所有相关事实的情形。然而，在许多情形中，我们并不知道这些事实。许多主观主义者主张，在其他这样的情形中

(M) 我们最有理由要做的事情，要能够最好地满足的欲望或目标，不是我们当下实际的目的式欲望或目标，而是如果我们知道并且合理地权衡了所有相关事实之后，我们现在会具有的或想要我们自己具有的那些欲望或目标。

许多人也会主张

(N) 当我们做出重大决策时，我们应当尽可能设法对我们行动的不同可能结果有更多的了解，以便能够获得更为充分知情的目的式欲望或目标，并能够由此设法满足这些欲望或实现这些目标。

我认为，主观主义者不能融贯地做出这样的主张。我们应当努力发现并且理性地权衡各种事实，这是因为这些事实可以给予我们某种理由。例如，陪审团应当考虑的是这样的事实，它们能够给予他们理由来相信嫌疑人是否犯有某种罪行。我们同样可以主张，当我们决定要努力产生何种结果时，在重要情形中我们应当努力发现并且合理地权衡，这些结果会是什么。但如果我们做出这样的主张，我们是在假定

(O) 这些可能的结果会具有其固有的某些特征，这些特征会给予我们对象给予的理由想要尽力产生或防止该结果。

(O) 就是客观主义者相信的东西。主观主义者否定(O)，按照他们的看法，这些结果从来就没有任何这样的特征，可以给予这种理由。如果情况真是这样，我们就没有任何理由要努力发现且合理地权衡，这些结果是什么。这样一来，这些人就不能融贯地断言(N)。

94

他们也不能融贯地断言(M)。如果正如主观主义者所主张的，(O) 是错误的，那么我们就没有任何理由相信：我们最有理由要做的是任何这样的行动，它能最好地满足的欲望或目标，不是我们当下实际的目的式欲望或目标，而是

我们理性地权衡了我们的行动所具不同结果的一切事实之后，我们现在会具有的那些欲望或目标。主观主义者不能称这些为给出理由的相关事实，因为这些人否认这些事实会给予我们理由。如果这些事实不能给予我们理由要具有这些欲望或目标，我们就没有理由要接受（M）。我们就没有理由认为，这些更知情的欲望或目标具有给出理由的更高地位，是我们更有理由要努力去满足的。

有些主观主义者做出了更弱的主张：

（P）我们有理由要满足的只是那些没有错误的当下目的式欲望或目标，意思是说这些欲望并没有依赖于错误的信念。

然而，这些人要捍卫这个主张也不得不诉诸（O），而这是主观主义者不能诉诸的。如果正如主观主义者所相信的那样，我们没有任何对象给予的理由，那么我们就没有任何理由想要更了解我们想要的东西，无论是获得新的真信念还是放弃当下的错误信念。

有些主观主义者承认，其理论具有这些含义。当科丝嘉捍卫“我们的合理选择使得某物为善”这种观点时，她说这种观点

使得我们在评估选择是否合理时，不用评估我们所选的事物：我们并没有必要确定特别合理的目的。

95 基于科丝嘉的观点，要理性地选择，我们没有必要评估所选之物的优缺点，因为任何事物都不具有优缺点，它们并不具有任何给出理由的特征。但大多数主观主义者没有看到：如果鉴于他们的假设，我们就没有任何理由要设法拥有并满足这种信息更充分的欲望或目标。如果正如我所论证过的，主观主义者无法诉诸（M）、（N）和（P），这就会破坏各种更精细且更可行的主观主义版本，诸如慎思理论和知情以及无错欲望理论。这些理论是不融贯的，因为它们假定下面两点：

（Q）我们的欲望、目标或选择给予我们理由，仅当在我们对想要之物的一切相关内在特征具有真信念的前提下，我们仍然会具有这些欲望、目标或做出这些选择。

且

(R) 这些特征不会给出任何理由让我们想要这些事物。

如果这些特征不给予我们任何这样的理由，这就会破坏这个主张：我们的欲望要给予我们理由，就必须是那种在我们拥有关于这些特征的真信念的情况下我们仍会持有的欲望。在此我们可称之为反对主观主义的不融贯性论证。我们可提到，这个论证完全独立于我先前的论证，因为这个反驳并不诉诸这样的主张：何种事实会给予我们理由。

不融贯性论证并不适用于下面这个更简单的目的式欲望理论，该理论的主张仅仅是：

(S) 我们最有理由做的事情，是能最好地满足我们当下实际的目的式欲望或目标的事情。

这种理论主张，无论我们的目的式欲望或目标是否依赖于错误的信念，我们都有这种理由。这些主观主义者可以融贯地主张：

(T) 我们应当努力发现有关“如何能最好地满足我们当下的目的式欲望或目标”的事实。

这些人之所以能够做出这样的主张，是因为(T)并不假定我们想要的事物或我们行动的可能后果具有给出理由的内在特征。基于目的式欲望理论，相关的事实并不包括有关结果是什么的事实，除非这些事实是有关“什么能最好地满足我们当下实际的欲望或目标”的事实。这些主观主义者还能融贯地主张：

96

(U) 如果我们想要有更知情的欲望或目标，我们就应当努力发现“有关不同可能结果是什么样子”的事实，以便能够具有这种欲望或目标。

然后这些人可以主张，既然大部分人确实想要具有这种更知情的欲望或目标，(U)就蕴含着大部分人应当努力获得它们。但是如同前面一样，这些主张

并不支持目的式欲望理论。我们大部分人之所以想要更知情的欲望和目标，是因为我们相信客观理论的主张。我们认为，我们行动的各种可能结果可能具有能给予我们理由的各种特征。

尽管目的式欲望理论并不是不融贯的，但它具有几种讲不通的含义，这导致许多主观主义者转向上面所讨论的其他主观理论。我的其他反驳是适用于目的式欲望理论的。基于这种理论，我们常常没有理由，想要过得幸福或避免未来痛苦，并且我们可能具有决定性的理由，为了其自身的缘故，致使自己处于痛苦之中，荒废我们的生命，努力去实现其他坏的和无价值的目标。

我已经论证过，不融贯性论证已经破坏了其他更精细且更可行版本的主观主义。还有一种更积极的方式来阐述该论证所揭示的东西。在诉诸“如果知道关于行动的可能结果之所有事实，我们会想要或选择什么”的主张之时，许多主观主义者正确地假定了这些结果可能具有给予理由的特征。例如，他们假定，我们拥有对象给予的理由想要趋乐避苦。这些人并不是真正的主观主义者。当这些人做出主观主义的主张时，他们并没有正确地陈述他们的真实信念。

我认为，法兰克福就是此类人之一。法兰克福写道，当我们决定关注什么时，我们并不需要理解什么是重要的。我们只要理解什么对我们是重要的就够了。为了阐明这个观点，法兰克福设想了一群人，他们的健康受到本底辐射的威胁。他写道，假定

97 某人真的完全不关注其健康……在这种情形下，本底辐射水平对他就是不重要的。它真的对他无关紧要；他没有任何理由要关注它。

法兰克福这里假定，如果我们不关注某事，我们就没有任何理由关注这件事情。基于这种观点，我们没必要理解什么是重要的，因为没有任何事物是重要的。真相仅仅是：某些事物对于某些人是重要的，这句话的涵义是这些人关注这些事物。

法兰克福给出了另一种阐述：

假定某人关注之事是避免踩上人行道的裂缝。毫无疑问他在关注此事上犯有错误……其错在关注并且因此给某个不值关注之事赋予了真正的重要性。

最后一句看起来蕴含着：有些事物是值得关注的，另一些则不值得。但法兰克福继续道：

它不值关注的理由看起来是很清楚的：“使得避免踩上人行道的裂缝对自己重要”对这个人来说并不重要。

如果“使得避免踩上人行道的裂缝对自己重要”对这个人来说是重要的，这句话就蕴含着：这个人就没犯任何错误。由此，避免踩上裂缝对这个人就是真正重要的，是值得关注的东西。法兰克福在其他地方主张，我们只要对事物有所关注，就由此回答了“如何生活的问题”。

法兰克福写道，人们可能会反驳：

人们实际上关注什么的经验论说……会错失我们“最初关注个人过何种生活这个问题”的全部意义。

法兰克福回应道：

98

不是有关关注的事实性问题错失了这意义，而是其规范问题。如果我们打算解决“在确定一种生活方式时的困难与犹豫不决”这一问题，我们在最根本上的需要的不是理由或证明，而是清晰与信心。

虽然是在论证我们应该拒绝规范性问题，但法兰克福设想：某个人首先决定他关注什么，然后

想知道他是否做对了……他变得关注这样的问题：他是否应该关注他事实上的确关注之事。

法兰克福主张，这个人的后一关注受到了误导。尽管不同的人关注不同的事物，但我们“并不需要决定谁是正确的”。法兰克福补充道，通过关注或爱某个事物，我们就是在给予我们的生活以意义，即使我们承认我们所爱的事物是坏的。

尽管最后一个主张是对的，但我认为，法兰克福的其他主张是错误的。法

兰克福写道

爱对于爱人者本身就是一种理由的根源。

我们可以同样主张

恨对于恨人者是一种理由的根源。

希特勒对犹太人的那种恨，给予了他人生最后一个阶段具有法兰克福所描述的那种意义。法兰克福所主张的那种（比具有理由更为重要的）“清晰性与信心”，希特勒是具有的。我认为，非常重要的一点是：希特勒是否有理由去做他所做过的事情，在我们的爱与恨之中，是我们还是希特勒做的事情是对的。

法兰克福捍卫其主张的方式，很让人迷惑不解。当某个人想知道他是否做对了事情时，法兰克福写道：

（V）“他在追问的是……是否可能存在更好的理由让他过另一种生活。”

99 如果我们追问的是这个规范性问题，法兰克福论证道，“我们注定会发现自己茫然不知所措”。任何试图回答这个问题的企图都不可能成功。有些人试图回答这类问题，因为他们认为某个目的或目标本身是好的，具有固有的（inherent）或内在的价值。这些人主张：

（W）当某个目的“具有固有的价值时……存在……某个理由选择它”，当某个目的“比任何其他事物具有更大的固有价值时”，这个理由是决定性的。

但是法兰克福写道，这个主张

（X）“并没有真正地处理，更不用说回答这个问题：如何恰当地确立个人的终极目的。即使这个主张是对的，它对人们如何选择他们要追求的目的还是根本没有提供任何论说”。

(X) 蕴含着，即使 (W) 是真的，(W) 也没有处理“我们要选择哪些目的”这个问题。如果主张 (V) 和 (W) 是在规范意义上使用“理由”，这个我们也可以“算作支持”来表达它，法兰克福的 (X) 就显然是错误的。如果某个事实决定性地算作支持我们选择某个目的，这就能帮助我们要选择那个目的。这表示法兰克福没有在规范意义上使用“理由”，那他为什么还要宣称，他正在讨论“我们是否具有更好的理由以其他方式来生活”这个规范问题呢？

我认为这个困惑是可以解决的。法兰克福在这些段落中错误地描述了他的真实观点。法兰克福还写道，当我们决定如何生活时，我们需要决定我们应该要努力实现几个可能目的中的哪一个。这些目的包括：

个人满足、快乐、荣誉、创造性、精神深度和与符合道德要求。

法兰克福的清单上并不包括：

个人的不满足、痛苦、玷污、无价值、精神上的浅薄和不道德。

鉴于法兰克福给出的、我们需要考虑的目的的清单，法兰克福似乎认为：某些目的在给出理由的意义上有内在价值，而另外一些则没有。法兰克福看来使用了“对象给予理由”的概念来持有某个目的或目标。事实似乎只能是，就像休谟和其他伟大的哲学家一样，法兰克福并没有真正意识到：他对此种理由的回应，影响了他有关我们应该关注什么和做什么的想法。

100

法兰克福在拒绝诉诸内在价值时写道：

哲学家之中似乎有种复兴的期望：存在某种终极的目的，对该目的的无条件采用，可揭示为某种意义上的理性要求。

就法兰克福的真实观点来说，存在许多内在地好的目的，但是没有哪个具有最高价值，也不存在“何种目的是最值得追求”的确切事实。我们常常必须在许多好的目的或目标之间做选择，其中没有任何一个显然比其他的更好，在此类情形中，理性并不要求我们选择任何一个目的。这些主张是说得通的，且完全不同于这种观点：没有任何目的就其自身而言是好的。我认为，法兰克福

并不是有关内在善性的虚无主义者，而是一个多元主义者。

法兰克福还正确地假定，理性可以要求我们采纳某些好的目的。法兰克福主张，我们需要具有目标，我们需要有充实意义的工作，这样我们的生活才不会空虚。我们有时候使自己具有值得追求的目标，从而给予我们的生活以意义。法兰克福假定，在这样的情形中，我们有决定性的理由采用某个这种目标，让自己的生活有意义。

101 当批判他所称的公然的理性主义理论时，法兰克福提出了这个重大的主张：我们并不需要理由来爱某个人。但是这个主张并不蕴含着，有如法兰克福有时候所表示的那样，爱是我们所有理由的根源。尽管我们并不需要理由来爱某个具体的人，但我们有理由尽量爱某些人，因为爱自身就是好的。爱以这种方式不同于恨。尽管爱是我们某些理由的根源，但这些通常也是对我们所爱之人做有益之事的理由，其途径是使得他们的生活按照我们和他们具有其他理由想要的生活方式来进行。正如法兰克福所提到的，在父母对其年幼子女之间的爱中，爱表现得最为简单和清楚。父母亲肯定有强有力的理由希望其孩子们做对的事情，采纳好的目的，并不特别关注（除非是出于一时兴趣）避免踩到人行道裂缝。

14 理由、动机和福祉

现在我们可以回到事件或结果之好坏的各种可能方式。对于两个可能的事件，一个可以

在蕴含不偏不倚理由的意义上是更好的，如果每个人从一个不偏不倚的角度来看都更有理由想要它或希望它会发生。

按照理由的主观理论，没有任何事件从不偏不倚的角度看会是所有人更有理由想要的，所以没有任何事件能在这种意义上比另一事件更好。例如，如果某个孩子的生命得救了，这不会是更好的。当任何孩子的生命得救时，还有很多人的知情欲望并没有得到更好的满足。即使每个人都有这个欲望，主观理论也不蕴含每个人都有理由具有这些欲望，即有理由想要任何孩子的生命得救。但这正是下面这个主张的意思：在蕴含不偏不倚理由的意义上，如果某个孩子的生命得救，情况会更好。

事件也可以是对特定的人更好的，意思是使得那些人的生活过得更好，或者更有助于其福祉。各种福祉理论可能在两方面相区别，因为福祉理论可以在两种不同的意义上使用“对…是好的”这一术语，并且可以在这些意义上对“什么对人们是好的”做出不同的主张。基于一切可行的福祉理论，每个人的福祉至少在部分程度上是过得幸福、避免痛苦。但对于其他东西对人们的好坏，不同的理论提出的主张存在部分的冲突。

当我们称某种可能的生活

在蕴含理由的意义上“对某人最好”的时候，我们意指这是这个人具有最强的自利性理由想要去过的生活，是其他人为了这个人的缘故具有最强的理由想要或期望他会去过的生活。

正如我说过的，“自利的”并不是意指“自私的”。即使是最利他的人也有理由关注自己的未来福祉。

如果我们接受理由的某种主观理论，我们就不能在这种蕴含理由的意义上使用“对某人最好”。主观理论蕴含着不存在任何自利性理由，因为这种理由是由未来事件的内在特征所提供的，这些特征会使得这些事件对我们或好或坏。主观主义者否认我们具有这种理由。

有些主观主义者认为，我们可以有不同类型的自利性理由。根据这些人的观点，既然我们大多数人关注自己的未来福祉，那么大多数人就有基于欲望的自利理由。这些主观主义者还主张，既然我们大部分人都关注道德，我们大部分人就都有基于欲望的道德理由。然而，基于这种观点，如果我们没有这些欲望，那么我们就没有这样的理由。在我虚构的情形1和2中，我将没有任何自利的理由来避免未来的痛苦。鉴于希特勒的欲望，希特勒可以没有任何道德理由不进行大屠杀。尽管主观主义者可以随心所欲地使用语词，但把这种基于欲望的理由称为自利理由或道德理由是误导性的。当我们大部分人使用这些语词时，除非某种理论蕴含着我们所有人都有自利的理由避免处于痛苦之中，否则这种理论就不会是有关自利理由的好理论；除非该理论蕴含着我们所有人都有道德理由不进行大屠杀，否则它也不会是有关道德理由的好理论。这样我们可以有证成地主张，按照主观理论来说，并不存在任何自利理由或道德理由。

对接受理由的主观理论的那些人而言，许多人是在不同于蕴含理由的意义上使用“对某人最好”。一个例子是罗尔斯在提出稀薄善理论（thin theory of

the good) 时给出的定义, 基于这种定义,

个人的善是由对他来说最合理的生活计划所决定的。

罗尔斯写道, 某种生活对某人最好, 只要这种生活能满足这样的规划——它是这个人

如果充分知情的话就会采用的。它对他而言是客观上理性的规划, 并决定了属于他的真正的善。

如果我们称某种生活

在基于当前选择的意义上“对某人最好”, 我们意指这是这个人在经过知情且程序上理性的慎思后会采用或选择的生活。

尽管这样的问题——何种慎思会是程序上理性的, 且在其他方面是理想的——是个规范性问题, 但个人在这种慎思之后事实上会选择什么却是一个心理学问题。基于这种观点, 理性就不会要求我们要具有任何目的式欲望或目标, 也许有些欲望是例外, 因为没有它们我们就无法慎思、选择要做什么和行动。罗尔斯写道, 对某人最合理的生活计划是这样的计划:

它是个人在具有充分的慎思理性后, 也就是说, 知道一切相关事实且对后果有过仔细的考量后, 本会选择计划。

无论我们具有什么样的目标, 我们都能在罗尔斯的意义上具有慎思理性。罗尔斯在其他地方主张, 从某个人是理想意义上理性的, 我们无法推出这个人确实(或会)想要什么或赞同什么。罗尔斯假定, 我们没有任何对象给予的理由想要把任何东西作为目标。

104 罗尔斯为了阐明他的善理论, 设想一个人选择的计划是以计算各种草坪上草的叶片数目来度日。罗尔斯根据他的理论写道, “这个人的善确实在于数草的叶片数”。罗尔斯假定, 这个人会享受以这种方式来过他的生活。但是根据罗尔斯的理论, 这个假设是没有必要的。只要在合理地考量了相关事实后, 这个人

事实上会选择这种生活计划，这就足够了。对于另一个例子，考虑

布鲁的选择：在这种理想的慎思之后，布鲁的最强欲望是其余生充斥无法减轻的痛苦。布鲁因此选择了会给予他这种生活的计划。

根据罗尔斯的理论，布鲁的最好生活就在于无法减轻的痛苦。

人们可能会主张这个例子太不现实，因为没人会选择一种无法减轻痛苦的生活。然而，正如我说过的，这种情形是否实际上出现是不相关的。罗尔斯并不认为，现实中有人会选择不计算草的叶片数目来度日，罗尔斯正确地把他的理论应用于虚构的人。任何可接受的规范性理论，必须有能力成功地应用于此类虚构情形。尽管很难相信有人会选择不痛苦无法被消除的生活，这恰恰是因为很难相信任何人会如此的不理性，以致选择在蕴含理由的意义上明显是如此坏的一种生活。然而，根据罗尔斯的理论，既然我们没有任何对象给予的理由，所以没有任何生活在这种意义上对某个人是坏的。用罗尔斯的话来说，“没有任何办法可以超越慎思理性”。

在某种意义上讲，我的例子完全没有反驳罗尔斯的善理论。当罗尔斯主张某种生活对某人最好，或真正有益于这个人的时候，他是在他提出的“基于当前选择的意义上”使用这些术语的。罗尔斯的意思是，这种生活是这个人合理的慎思之后实际上会选择的。因此，罗尔斯是正确地提出如下主张：在他提出的意义上，这是对布鲁最好的生活。这不过是换一种方式来说：这种生活是布鲁在理性慎思之后会选择的生活。

然而，罗尔斯的主张想做的并不止是这点。罗尔斯提出的这种意义上的“对某人最好”是想代替这个术语的日常意义，从而以一种更清楚的方式来说我们想要说的一切。我认为，罗尔斯想要说的是：如果布鲁的生活不具有无法减轻的痛苦，那对他会更好。

罗尔斯如果在其他某种意义上使用“对某人最好”，那就可以做出上述主张。既然罗尔斯是理由的主观主义者，他就不可能在蕴含理由的意义上使用“对某人是最好的”。但这个术语通常是在其他意义上使用的。当人们称某种可能的生活“对某人最好”时候，其中有些人的意思

这是这个人在减去痛苦后具有最大数量幸福的一种可能生活。

其他人的意思是

这是这个人在不同时间段的欲望得到最好满足的一种可能生活。

我们可以称此“对某人最好”这个术语的享乐主义意义和时间中立的、基于欲望的意义。罗尔斯确实能够主张，在这些意义上，布鲁的生活具有无法减轻的痛苦对他来说是坏的。在享乐式意义上，这种生活对布鲁来说是极坏的。尽管这种生活能够最好地满足布鲁在选择之时的欲望，但其余生之中的其他欲望则远远得不到满足。

然而，下述主张意义不大：这种生活在享乐式意义上对布鲁是不好的。这个主张在享乐主义意义上讲则是另一种隐藏的反复，其公开的形式是这种无足轻重的主张：如果布鲁的生活包含更多的痛苦，它就将包含更多的痛苦。在时间中立且基于欲望的意义上，这个主张也是相当无关紧要的，因为它的意义仅仅是：如果布鲁的生活包含更多的痛苦，他的欲望就满足得不好。这也同样适用于其他情形。当人们在这些意义上使用“对某人最好”时，他们不可能对于“何种生活对人们最好”具有实质性的规范信念。

如果这些人接受理由的某种客观理论，他们就能具有这种实质性的规范信念，这样他们也能在蕴含理由的意义上使用“对某人最好”。他们因此可以主张

106

(V) 如果某种可能的生活在享乐主义的意义和时间中立且基于欲望的意义上对某人都是最好的，这些事实会使得这种生活在蕴含理由的意义上对这个人是最好的。

这意指

(W) 如果某种可能的生活既能给予某个人最多的幸福，这个人的欲望在整体上还得到最多满足，这些事实会使得这个人具有最强的自利理由想要和努力去过这种生活，这也是其他人为了这个人的缘故而有最强的理由想要或期待他会去过的生活。

这个主张是实质性的且是可行的。但是如果我们接受理由的某种主观理论，那么我们无法做出这样的主张。

理由的主观主义者也许会在其他意义上使用“对某人最好”，但这无助于他们避免其结论的不可行性。经过理想的慎思，布鲁最强的欲望及其已选目标就是一种无法减轻痛苦的生活。主观理论不可避免地蕴含

(X) 即使一种无法减轻痛苦的生活在其他意义上对布鲁是坏的，但这是布鲁现在最有理由尽力想给自己的生活。

如果布鲁现在能够确保他具有这样的一种生活——让他自己成为某个残酷主人的奴隶，或者因犯罪受罚而去做无尽的苦力，那么基于主观理论，这将是布鲁最有理由要去做的事情。如果他知道这些事实的话，这是他应当合理地去做的。

这也同样适用于实际的情形。主观理论蕴含着：我们没有任何对象给予的理由想要自己或他人过幸福的生活，也没有任何这样的理由具有任何其他好的目标。正如我论证过的，主观主义不能有证成地主张：我们具有主体给予的理由要具有这样的目标，或者为了事物自身的缘故而关注任何事物。这种理由必须为我们根本没有任何理由要具有的某个欲望或目标所提供，而这样的欲望或目标不可能有证成地主张给予我们任何理由。这样一来，我们现在能够断定：基于这些被广泛接受的观点，没有任何事情是重要的。

107

有些主观主义者会承认，基于他们的观点，没有任何事情在非个人的意义上是重要的。这些作者宣称：某些事情对于某些特定的人是重要的，这就足够了。而这个回答表明，我们一直在考量的两类理论之间的分歧有多深。按照客观理论，有些事情是在规范意义上重要的，意思是指我们有理由要关注这些事情。当主观主义者主张某些事情对于某些特定的人是重要的时候，他们仅仅意指这些人确实关注这些事情。这不是一个规范性主张，仅仅是一个心理学主张。我们都知道人们关注某些事情。我们期待哲学家或其他的智者告诉我们更多的东西。

主观理论除了蕴含没有任何事情重要之外，甚至都不可能证成地主张我们有任何的行动理由。正如我论证过的，我们的欲望、目标和选择无法证成地主张给予我们任何这样的理由。

15 支持主观主义的论证

这些黯淡的观点很少得到捍卫。大部分主观主义者理所当然地认为，理由

是由有关我们的欲望或目标的事实所提供的。

在主观理论的辩护者之中，有些人诉诸“‘应当’蕴含着‘能够’”这一主张的某种版本。这些人如此论证：

(1) 我们要有理由去做某事，我们必须确实能够做该事。

(2) 如果即使在理想的慎思后，还是不想或没有动机做这件事情，那我们就没能力去做它。

108 因此，

我们要有理由去做某件事情，情况必须是：在这种慎思之后，我们会有动机去做它。

但在相关的意义上情况并不是(2)这样的。假定我说，“你本该帮助那位盲人过街”，而你说，“我不可能那样做”。如果我问，“为什么不能呢？”，那么“因为我不想”这样的回答是不够的。某些特殊情况除外，如果除了“我不想做某件事情”外没有其他任何事实阻止我们去做，那么我们在相关的意义上就是能做这件事情的。

有些主观主义者的论证是：

(3) 我们若具有规范性理由，就可以出于这个理由而行动。

(4) 我们若出于这个理由而行动，就有动机这样做。

(5) 既然我们有动机这样做，这个理由就是基于欲望的。

所有的行动理由都是基于欲望的。

因此，

但是(5)是错误的。我们无法证成如下主张：只要人们的动机是某种理由，这种理由就不能是对象给予和基于价值的，而必定是主体给予和基于欲望的。这个主张要得到证成就不得不假定：对于对象给予和基于价值的理由，任何人都不可能有机出于这种理由而行动。这个假设是荒谬的。如果某个行动能实现某个好的或有价值的目标，我们很可能会出于这种理由而行动。

这些主观主义者也许会回应道：

(6) 只要我们行动，我们就有动机这样做，因此我们总是有某种基于欲望的理由去做我们所做的。

因此

(7) 所有的行动理由都是基于欲望的，即使可以主张其中某些理由也是基于价值的。

因此

在我们的实践理由的论说中，诉诸基于欲望的主观理论就足够了。

但是(6)要么是搞混了规范性理由与驱动性理由，要么是主张，只要我们行动，我们因此就给予自己一个规范性理由去做我们所做的事情。后一主张错误地假定：不管一个行动多么疯狂，它都能在部分程度上自我证成。当这个论证认为(6)蕴含(7)时，它也错误地假定：只要我们没基于某种理由而行动，我们就不可能具有这种理由。(7)则错误地假定基于价值的理由也是基于欲望的。

还有另一种更重要的思路导致人们成为主观主义者。这些人做出了某些元伦理假设，我会在第六部分讨论这些假设，这里仅简要提及。基于最好的客观理论，“我们具有某理由”这一事实是一个不可还原的规范真理。很多接受主观理论的人都是形而上学上的自然主义者，他们认为不可能有这样的事实或真理。按照这些自然主义者的看法，所有的属性和事实都必须是能由自然科学和社会科学加以探究的。这些人假定，不可还原的规范真理与科学的世界观是不相容的。

大部分自然主义者都接受这样的论说：行动理由是基于欲望或目标且是还原性的。根据某些分析性主观主义者的观点，若我们主张某人有理由以某种方式行动，则我们意指这个行动能满足这个人的目的式欲望，或者是在知情的慎思后，这个人会选择这样做，或者意指其他的此类情形。按照其他某些自然主义者来说，尽管理由的概念是不可还原地规范的，但是“某个人具有理由”这个事实是或在于某种上述因果或心理学事实。

如果像许多人一样，我们把任何理由的规范性或规范效力看作某类驱动力，那么这些还原性的主观理论可能看起来是说得通的。我们可能因此会认为，我们应该把行动理由等同于这样的事实：什么会满足我们当下的欲望或我们如何受驱动去行动。就解释这些理由的规范性而言，这就可能看起来是最好甚至是唯一的方式，如同形而上学意义上的自然主义者所做的那样。正如某些这样的人所写的：

哲学自然主义者关注把规范性放在自然秩序之内，对他们来说，规范效力要想不是驱动性力量，这是完全说不通的……

基于最深刻的反思，价值若成为完全不同于打动或能打动评价者（有些事物对他们能够是重要的行动者）的东西，看来它就什么也不是。

我们不可能如此来看待对象给予的、基于价值的理由，因为我们即使没有被打动或受驱动而据之行动，还是具有这样的理由。

那些给出这种还原性论说的人，大部分都主张是在描述规范性理由。但我认为，按照这些观点，不可能存在任何真正的规范性理由，只存在行为的原因。事物只有在如下意义上才是重要的：某些人关注这些事物，这些关注能打动这些人去行动。

我认为，理由的这种自然主义论说存在深刻的错误。我在第六部分将捍卫这一信念，而我在此只说一句：如果形而上学的自然主义为真，我们就不可能有理由具有任何特定信念。这种认知理由也是不可还原地规范性的，因此也要受到相同的自然主义反驳。由此情况不可能是我们应当接受自然主义，我们也不能具有任何理由要接受这种观点。我们要想能合理地论证自然主义是否为真，自然主义就必须是错的。

我认为，自然主义是错的。有些事情在完全不同的意义上是重要的，我们有理由要去关注这些事情。

第五章 合理性

16 实践合理性与认知合理性

我们现在可以从理由转向合理性。正如我说过的，我们若意识到各种事实给予我们某些理由时，就应当理性地对这些理由做出回应。有些事实给予我们决定性的理由相信、想要或去做某些事情，若意识到这一点就会导致我们相应地行动，我们就对决定性理由做出了回应。只要我们没有这样来回应，我们就是不理性的，或者不是完全理性的。但只有我们意识到某些事实给予了这种理由时，我们才能在回应这种理由上失败。 111

尽管理由是由事实所给予的，但我们能理性地想要或做的事情依赖于我们的信念。如果我们对“给予理由的相关事实”具有某种信念，并且这种信念的内容（若为真）会给予我们某种理由，我就称这些信念为真就给予我们这种理由。这样的信念给予我们显见理由。当这种信念为真时，这种显见理由就是真实理由。这些信念包括我们没有明确地意识到的假设，诸如“某个行动不会伤己或害人”这种假设。我们若无知或具有错误的信念，就可能想要或去做没有任何理由想要或去做的事情时，但我们仍可能是理性的。由此，我们回应的只是看起来是理由的因素。即使因为我们具有错误的信念，由此显见理由不是真实理由，我们也应当对这些理由做出合理的回应。

接下来我们可以更细致地考察，欲望与行动的合理性如何依赖于信念。我有关欲望的主张也适用于目标。我们由于持有如此信念而拥有如此欲望并依此行动，当此之时，我们的欲望和行动就因果地依赖于我们的信念。某个欲望可能会因果地依赖于某个完全不相关的信念。我们能想象，我想要去睡觉是因为7是个质数。但是如果我的欲望是直接地依赖于这种信念，我可能是精神上 112

有病，或具有某类局部的脑损伤。7是质数不能给我任何理由想要去睡觉。在大部分情形中，当某个欲望依赖于某个信念时，这种关系不仅仅是因果性的。我想要去睡觉可能是因为我相信：我除非睡觉，否则明天的面试会表现很差。既然这个欲望是对我所相信之事的合理回应，这个欲望就不仅仅是由我的信念所引起的，而且是由它所证成的。我现在将简要地捍卫关于如下问题的一种观点：我们的欲望和行动如何能或不能为我们的信念所证成。

我们某些欲望的合理性仅仅依赖于其意向性对象，这个对象是我们想要的可能事件，且带有我们认为它们会具有的特征。当我们想要之事的特征能够给予我们想要此事的理由时，这种欲望就是合理的。例如，我们想要避免痛苦就永远都是合理的。我们其他欲望的合理性在部分程度上依赖于有关我们想要之物的其他信念。例如，我们想要既安全又能减轻痛苦的药物，这种欲望就是合理的。这也同样适用于我们的行动。我们行动的合理性既依赖于我们有意去做的事情，还依赖于我们正在做什么的其他信念。基于这种观点：

113

(c) 若我们的欲望和行动以正确的方式因果地依赖于某些信念，且这些信念若为真会给予我们充分的理由具有这些欲望且依此行动，则这样的欲望和行动就是合理的。

我们可以补上

(d) 在大部分情形中，这些信念是否为真或合理是不相关的，例外情形涉及到某些规范性信念。

(e) 如果我们的信念是不一致的，那么我们的某些欲望或行动可能相对于我们的某些信念是合理的，而相对于另一些信念则是不合理的。当我们对于给予理由的相关事实没有任何信念时，那么没有任何事情是我们应当理性地去做的。

(f) 当且因为某个欲望本身是合理的时候，我们具有这种欲望就在某种意义上是合理的。但是在某些情形下我们能理性地让我们具有某些不合理的欲望。那么我们具有这种欲望就在另一种意义上是合理的。我们让自己不理性地行动就可能是理性的。我在附录A和B中会进一步探讨这种情形。

我们若要是完全理性的，就需要满足某些理性要求，诸如没有相矛盾的意图、打算做我们认为应当做的事情之类的要求。我在此不会探讨这些要求。

许多人会拒绝某些这样的主张。休谟表示，我们的欲望仅当因果地依赖于错误的信念时才是不合理的。但是错误的信念可以是合理的，依赖于错误信念的欲望亦如此。

基于一种更广泛持有的观点，我们的欲望只有当因果性地依赖于不合理的信念时才是不合理的。为了评估这种观点，我们可以假设：我想抽烟是因为我想保持身体健康，且相信抽烟是达成这个目标的最有效手段。我之所以具有这个不合理的信念，是因为我抽烟的邻居活到了100岁，并且我认为这个事实压倒了抽烟致命的一切证据。为简化起见，可补充一点：我并不喜欢抽烟。我之所以想要抽烟，仅仅是因为我喜欢活着且相信抽烟能延长我的寿命。我不合理的信念会使得我抽烟的欲望成为不合理的吗？

我认为，最好对此予以否定的回答。使得我们的欲望合理与否的东西，不是“该欲望因果地依赖的那些信念”的合理性，而是这些信念的内容，或者说我们相信什么。鉴于我的信念是抽烟会有益健康，我抽烟的欲望就是合理的。如果我的信念为真，我想要的东西就是我有很强理由想要的。相反假定我想要抽烟是因为我们具有这个合理的信念：抽烟有害身体。基于我们现在讨论的观点，既然我抽烟的欲望在此依赖的是合理的信念，这个欲望就是合理的。这显然是错误的。我因为相信抽烟有害身体而想要抽烟，这显然是不理性的。

114

我们接下来假定某个修士想要过一种苦行僧生活，因为他不理性地认为，这样做会取悦上帝。这个修士的欲望鉴于其信念就是合理的。如果这个修士想要过这样的生活是因为他相信这样做不会取悦上帝，他的欲望就是不合理的。

这样的主张也适用于我们的行动。在大部分情形中，如果行动所依赖的信念为真就会给予我们充分的理由这样做，那么这种行动就是合理的。既然我具有“抽烟有益健康”这个不合理的信念，我抽烟就是合理的。既然那个修士具有“苦修式生活会取悦上帝”这个不合理的信念，他这样活着就是理性的。我们的主张只能是这样的：既然这些不合理的信念是错误的，我和修士就没有任何理由那样去做。

有些人可能会做如下反驳：他们若称某个欲望或行动是“不合理的”，就是指这个欲望或行动因果地依赖于某个不合理的信念。如果这些人的意思就是这样，那么我无法否定他们的主张：欲望和行动若依赖于错误的信念就是不合理的。但我认为，我们应当在日常意义上使用“不合理的”来表达我们也用

“笨的”、“愚蠢的”和“无意义的”等所做的那类强烈批评。我提议，我们应当对何种欲望或行动应得这种批评做出不同的主张。

有些人主张欲望的合理性依赖于信念的合理性，其中许多人都认为我们没有任何理由要具有欲望。这些人假定，欲望只有在“因果地依赖于合理或不合理的信念”这种派生意义上，才可能是合理的或不合理的。但是我们确实有理由具有某些欲望。如客观主义者主张，我们有理由想要某些作为目标的事件；而主观主义者也主张，我们通常有理由想要作为实现目的或目标之手段的事物。既然我们可以有理由拥有我们的欲望，那就应该主张：欲望的合理性取决于我们在具有这些欲望时是否对这些理由或显见理由做出了恰当的回。我们还应该主张，如果我想要抽烟，那么我处于不理性之状态，但是这种不理性不在于我的欲望而在于我的信念。

我们还有其他的理由拒绝这样的观点：我们的欲望或行动仅当依赖于不合理的信念时才是不合理的。这种观点即使只应用于信念也过于狭隘。假定因为我相信下面两点：

(1) 抽烟有益健康。

(2) 我现在在抽烟。

我认为

(3) 我现在正保护健康。

我的信念(3)可能在某种意义上是不合理的，因为这种信念在部分程度上依赖于我不合理的信念(1)。然而，我的信念(3)在另一种意义上是合理的。这个信念是合理地派生于我的信念(1)和(2)，意思是，如果这些信念是对的，这就给予我决定性的理由相信(3)。给定我的信念(1)和(2)，如果我现在问自己这个问题，那么我不相信自己正在保护自己的健康就在某种意义上是不合理的。我们因此主张

(E) 某个信念是否合理，在部分程度上依赖于这个信念是否合理地派生于我们的其他某些信念，还在部分程度上依赖于其他的这些信念是否合理。

我们某些信念的合理性在部分程度上依赖于其他事情，比如它们与我们的感知经验的关系。但是应用于我们的许多信念时，(E)大致是正确的。

我们也可以对欲望和行动做出类似的主张。我们通常具有某个欲望或以某种方式行动，因为我们具有这样的信念——其为真给予我们充分的理由拥有该欲望或这么做。我们可以称这种欲望和行动受到这些理由的合理支持。我们可能会提出

(F) 某个欲望或行动是否合理，在部分程度上依赖于它们是否受到我们某些信念的合理支持，还在部分程度上依赖于这些信念是否合理。

这里改变一下我的例子，假定我想要去某个拥挤且吵闹的派对，因为我相信自己会喜欢它。既然我到现在应当已经搞清楚我从来不喜欢这样的派对，这个信念就是不合理的。基于(F)表达的观点，鉴于我的信念是不合理的，我想参加这种派对的欲望在一个方面就是不合理的。然而，我的欲望在另一个方面是合理的，因为它受到我的信念的合理支持。我想要自认为喜欢的东西，这是合理的。如果我之所以想去参加这个派对，是因为我具有我不喜欢它这个合理的信念，我的欲望会在某种意义上是不合理的。

我们接下来假定，格林做某件事情是因为她具有这个不合理信念：这个行为肯定能实现她的目标。格蕾做某件事情是因为她具有这个不合理信念：这个行动肯定会挫败她的目标。按照(F)来说，格林和格蕾的行动都是不合理的，因为她们的行动都依赖于不合理的信念。但是在格林的行动在另一种意义上是合理的，而格蕾的不是，因为我们去做自认会实现自己目标的事情是合理的，而去做我们认为会挫败自己目标的事情是不合理的。

尽管(F)是讲得通的，但我认为这个观点并不是最好的。按照(F)来说，当且因为我们的欲望或行动没有成功地回应某个认知理由或显见理由，它们在某个方面才是不合理的。当我抽烟是因为我认为这有益健康时，我的行动在这个方面是不合理的。但我们称我的行动是实践上不合理的，这是误导性的，因为我的错误仅仅是没有成功地回应“不要有这种信念”的认知理由。我们称这个行动是认知上不合理的，这也是误导性的，因为并不是由于我这么做，我才没能回应这些认知理由。

我提议，我们不应该做出这些误导性的主张。当某个信念是认知上不合理的时候，可以言之有理地做出这样的有益主张：这种不理性可以传递给依赖于该信念的其他任何信念。但如下的主张却是没有价值的：某信念的不理性也会传递给依赖于该信念的任何欲望或行动。既然认知理由和实践理由之间具有这种差别，我们就应该转向另一种更简单的观点。我们应该主张，只有信念才能是认知上不合理的。我们可以用另一种比喻的说法：当某个信念是认知上不合理的时候，这种不理性就像病毒一样可以传染给其他的信念。但是除了我马上就会谈到的几个例外，这种不理性无法穿越信念与“欲望和行动”之间的那道鸿沟。我们的欲望和行动，仅当未能回应清晰且强决定性的实践理由或显见理由“要不具有这种欲望或这样行动”时，才最适合被称作是不合理的。

基于这种更简单的观点，信念的合理性依赖于这一点：我们具有这些信念时是否很好地回应了“要具有这些信念”的认知理由、真值相关的理由或显见理由。我们的欲望或行动的合理性依赖于这一点：我们具有这些欲望或这样行动时是否很好地回应了“要具有这些欲望和那样行动”的实践理由或显见理由。我们可能对其中一组理由或显见理由的回应很好，但对另一组理由的回应则很差。我们可能是实践上理性的但认知上不理性的，或者实践上不理性的但认知上理性的。

我们接下来能简要地考虑另一种广泛持有的观点。基于这种观点，认知合理性的独特性在于要获得真信念这个目标。当我们以我们认为会最好地实现这种认知目标的方式行动时，我们在认知上是理性的，并且回应了认知理由。尽管我们无法主张这个观点是错误的，但我认为，它不是最好的观点。除了更清楚地区分认知合理性与实践合理性，我们最好是以不同方式和在不同的地方来做出这个区分。这里最深刻的区分不在于

获得真信念的目标和其他可能目标。

当我们以自认会最好地实现我们的合理目标的方式行动时，我们在实践上是理性的；并且不管这个目标是什么，我们都回应了实践的理由或显见理由。这个深刻的区分在于

我们对实践理由的回应是自愿的，而对认知理由的回应是非自愿的。

努力获得真理是一种活动，我们是出于实践理由而从事它。例如，我们在运算时可以有实践理由来检查某个证明，或者以另一种方式重新计算，确证早前的计算结果。我们在这样做来回应这些实践理由的同时，也以非自愿且更直接的方式在回应许多认知理由。例如，当我们在检查某个证明时，只要我们相信，既然某事是真的，其他某事也必定是真的，这就是在回应认知理由。我们具有这种具体的信念并不是一个自愿的行动。正如我在附录A中所表示的，实践理由和认知理由支持的是对不同问题的答案，所以它们是不可能相互冲突的。

17 有关理由的信念

在没有有关理由的任何信念的情况下，我们也能具有合理的信念和欲望，并且合理地行动。尽管小孩子没有任何理由的概念，他们还是能合理地回应某些理由或显见理由。猫、狗还有其他某些动物尽管从来都没有理由这个概念，但还是能对某些理由（诸如相信我们要喂养它们）做出回应。某些理性的成年人似乎缺乏理由概念，或者忘记他们具有这个概念。例如，休谟宣称，任何欲望或偏好都不可能是不合情理的，此时他似乎忘记了这个概念。

如果我们有“何种事实给予我们理由”的信念，我们的欲望和行动就通常是对我们信念内容的合理回应。但情况并不总是如此。我们大部分人想要的某些事物，是我们根本没有理由想要或有强有力理由不想要的。当许多精疲力尽的父母想要揍他们号啕大哭的孩子时，情况就是这样；当我想砸烂某些用不了的机器时，我也是处于这种情况。当我们具有的某些欲望是我们没有理由要有且有理由不要有的时候，我们具有这种欲望就不是完全理性的。可以说，这类欲望与我们的规范信念是不一致的或说不匹配的。

我一直主张，在大部分情形下，如果欲望所依赖的信念若为真就会给予我们充分理由持有这些欲望，那么这样的欲望就是合理的。我还主张，在这种情形下，这些信念是否为真或合理是不相关的。不过，我们的欲望若在部分程度上依赖于我们的某些规范信念，则这个主张就不适用，这些信念是否为真或合理就可能是相关的。假定我们错误且不理性地相信：某个事实给予我们理由具有某个欲望，且这个欲望是合理的。如果这些信念为真，我们就有理由具有这个欲望，并且这个欲望会是合理的。但这并不使得我们实际上真有这样一个理由，也不能使得这个欲望是合理的。这也同样适用于行动。如果我们错误且不理性地相信，我们有理由以某种方式行动或者某个行动是合理的，这并不给予我们

119

这样的理由，也不会使得该行动是合理的。实践合理性没有这样容易达到。

有人可能会反驳道，当我们对“何种事实给予我们理由”具有不合理的信念时，这并不使得我们在实践上不理性。既然这些是信念，我们就处于认知上不理性的状态，因为我们没有成功地回应我们的认知理由从而不具有这些信念。而正如我一直主张的，实践合理性与认知合理性是极为不同的。

然而，正如以前一样，这个主张也仅适用于大部分情形。当我们的信念是有关实践理由时，这些类型的合理性与理由就重叠了。正如斯坎伦提到的，我们的许多欲望都可以被更充分地描述为受（某事在蕴含理由的意义上是好的或值得实现的）信念驱动的状态。既然这些欲望和信念之间的关系极为紧密，这些欲望的合理性确实在部分程度上依赖于这些信念的合理性。我们具有的不合理信念如果是关于实践理由以及我们应当合理地想要或想做什么，那么就使得我们以某种方式是实践上不合理的。

120 实践理由和某些认知理由也有类似的重叠。例如，我们具有实践理由想要避免处于痛苦之中，也具有认知理由相信我们具有这种实践理由。痛苦的性质既给予我们这种实践理由，也使得“我们具有这种实践理由”显然为真，从而也给予我们这种认知理由。

我已经说过，无需对“何种事实给予我们理由”具有任何信念，我们的欲望和行动也可以是合理的。只要我们理性地对“意识到的、给予理由的事实”做出回应就够了，或者我们的行动是基于非规范性的事实信念，且其为真就会给予我们理由，这也够了。但是当我们具有“何种事实给予我们理由”的信念时，仅当这些信念是合理的，且仅当我们也想要、意图和努力去做我们具有决定性理由想要、意图和努力去做的任何事情，我们才是实践上完全理性的。

按照某些作者的看法，我们并不需要对理由或显见理由做出很好的回应才能是完全理性的。我们只需要满足某些合理性要求，诸如想要和意图“我们相信我们具有决定性理由想要或意图”的任何事情，这就足够了。我认为，这样的观点过于狭隘。

为了阐明这种分歧，假定

斯加力宁愿明天剧痛一小时，也不愿要下周其他时间微痛一分钟；
克里蒙森宁愿明天剧痛一小时，也不愿今天随后微痛一分钟；

以及

平克宁愿明天微痛6分钟，也不愿今天随后微痛5分钟。

对剧痛和微痛、个人同一性、时间，对于其他一切相关的非规范性事实，这些人具有的信念都为真。但是这些人对于理由的信念则有所不同。

斯加力的观点是我们前面接触过的。按照他的观点，除了任何一个未来的周二外，我们有理由关心我们未来的一切遭遇。既然明天是周二，斯加力就认为：他有决定性的理由宁愿明天剧痛一小时，也不愿其他时间稍痛一分钟。斯加力具有这种偏好，所以他选择这种痛苦。

克里蒙森的观点更为接近现实中许多人所接受的观点，他认为，尽管我们121有理由关注一切未来，但是我们关注更近未来的理由更为有力。因此，克里蒙森认为他有决定性的理由偏爱明天剧痛一小时而不是今天随后微痛一分钟。克里蒙森具有这种偏好，因此他选择具有这种痛苦。

基于平克的观点，既然单纯时间上的差别没有任何理性意义，因此我们应当同等偏爱一切未来。平克由此认为，他具有尽管很弱却是决定性的理由偏爱今天微痛5分钟而不是明天微痛6分钟。尽管具有这种信念，但他还是偏好并且选择明天稍长时间的微痛。

当斯坎伦讨论某个具有斯加力式偏好的人时，他写道，“这样的一个人不是不理性的，他只是极为错误的。”斯坎伦建议，只有当某个人“没有成功地回应他认为是相关理由的因素”时，我们才应该称他为不理性的。

如果斯坎伦是在日常意义上使用“不理性的”这个词，那我认为，他的主张是没道理的。斯加力避免了一种不理性，因为他的偏好与其有关理由的信念是相匹配的。但是他没有关注他的未来痛苦，由此没有回应非常清楚且强有力的理由。尽管他的偏好匹配其规范信念，但这个信念是极不合理的。我们有理由避免痛苦，但周二的痛苦除外，这种信念是疯狂的。我认为，这些事实足以使得斯加力的偏好是不合理的。

既然克里蒙森的偏好并没有随意划出界线，并且认为我们有理由更为关注更近的未来不是讲不通的，所以他的偏好没有那么不合理。但是这种观点的克里蒙森版本太过极端了。我们具有决定性的理由宁愿明天剧痛一小时，也不愿今天随后微痛一分钟——这种信念是不合理的。既然克里蒙森的偏好匹配他有关理由的信念，所以他也避免了一种不理性。但在宁愿剧痛而不是微痛时，他没有回应一种清楚且强有力的决定性理由；而且他的偏好和信念之所以相匹

配，仅仅是因为它们都是不合理的。

122 由于平克的偏好并不匹配于其有关理由的信念，所以他在某一方面没有斯加力和克里蒙森那么理性。但我认为，这个事实被其他两个事实压倒。就其偏好的拥有而言，平克未予回应的理由要弱得多。斯加力和克里蒙森宁愿多要一个小时的剧痛，而平克宁愿要的只是多一分钟的微痛。不同于斯加力和克里蒙森，平克具有一个合理的理由信念。我认为，这些事实使得平克不理性的程度在三个人中是最轻的。

斯坎伦主张，在人们未予回应其自认是理由的东西之时，他们的不理性最为明显。这个主张在某种意义上讲是对的，因为即使是依据其自身的信念来看，这些人也不是完全理性的。如果指控这些人不是完全理性的，他们得认罪。但这并不能证成这个主张：只有这样的人才能被称为不理性的。基于斯坎伦的观点，即使我们常常不能回应清楚的决定性理由，但只要我们对于“何种事实会给予我们理由，何种欲望或行动是合理的”不具有任何信念或持有错误的信念，我们就可以避免不理性。我认为我们应当拒绝这种观点。斯加力对于未来周二的态度就是不合理的，即使他认为它是合理的。如果我们对于实践理由具有合理的信念，且我们承认但没对这些理由做出回应，而有些人对实践理由持有不合理的信念，且他们未承认的错误要严重得多，那么我们不理性的程度并没有这些人那么严重。

这种主张也适用于信念。按照与我类似的观点，如果我们具有清楚且强有力的决定性认知理由要不持有某些信念，那么我们持有这些信念就是不合理的。基于某种斯坎伦式观点，只有当我们没有回应自认为是相关的理由时，我们的信念才是不合理的。假定我尽管知道赢取彩票的概率只有十亿分之一，但认为这个事实并不给予任何理由让我放弃“我会赢”这个信念。我尽管知道没有任何其他人能徒手与十只狮子搏斗而活下来，但认为这个事实没有给予我任何理由放弃“我能从这样的搏斗中活下来”这个信念。基于一种斯坎伦式观点，这些信念并非不合理的，因为我只是在关于何种事实给我理由这个问题上犯下大错。然而，我在具有这些信念时并没有回应清晰的强决定性理由，这就足够使得这些信念是不合理的。

123 我们的欲望和行动仅当与我们的规范信念不匹配时才是不合理的，这种观点还有另一个版本。按照某些人的说法，既然关于“理由或什么是合理的”没有任何真理，由此仅当我们自己相信我们是不理性的时候，我们才是不理性的。许多人对道德提出了这样的主张。按照这些人的说法，既然没有任何道德

真理，每个人就都应做他们自认为应做的任何事，而任何人都会行动不当，除非他们在做自认为不当之事。这里的道德怀疑论导致了一种不一致且自我破坏的相对主义。

我们大部分人正确地拒绝了这样的观点。如果我违背的诺言是无足轻重的，或者讲的谎话是无伤大雅的，那么尽管我相信这样做是不当的，但我的行动只是稍有不当。但当某个党卫军军官杀死大量市民并认为这是其义务的时候，他的行动就极为不当。与我不同，这个人认为他的行动并非不当的——这可以是某种辩护，但他的行动在道德上比我的行动要坏得多。我认为，当我们讨论合理性时，这样的主张也是适用的。在我虚构的人之中，只有平克没有回应他认为是理由的因素。但是斯加力和克里蒙森是不理性的，而平克仅仅是不完全理性的。

我已经拒绝了斯坎伦的这个主张：当像斯加力和克里蒙森的人宁愿要一个小时的剧痛而不是一分钟的微痛时，这些人的偏好不是不合理的。然而，我们在此也可能没有任何分歧。我是在日常意义上使用“不合理的”，意思大概是“应该得到如同‘很笨的’，‘愚蠢的’和‘疯狂的’等词所表达的那类强烈批评”。而斯坎伦在某种意义上是建议我们应该在他所称的“较狭窄的”意义上使用“不合理的”，这种意义仅仅适用于人们没有回应他们自认为是理由的因素，或者以其他方式是不一致的。如果斯坎伦是在这种狭窄意义上使用“不合理的”，他与我的观点并不冲突。我同意，当斯加力宁愿要一个小时的剧痛而不是一分钟的微痛时，他在这种意义上不是不理性的。斯坎伦可能同意：斯加力犯了一个极大的实质性错误，并且相对于平克宁愿多要一分钟的微痛来说，斯加力宁愿多要一个小时的剧痛应得更强烈的批评。然而，如果这就是斯坎伦的观点，那么他说只有平克的偏好是不合理的，这就是误导性的，因为那会暗示平克的偏好应得更强烈的批评。我认为我们应当在其日常意义，也是更宽泛的意义上使用“不合理的”。如果我们认为两种偏好之一在合理性上应得更强的批评，我们就不应该说只有另一种偏好是不合理的。

124

我们接下来可以简要地查看一下不同版本的虚构情形。我们现在可假定斯加力和克里蒙森都是理由的主观主义者。这些人尽管具有上述偏好，但认为他们没有任何理由持有这些偏好的。按照他们的观点，我们没有任何理由想要把任何事物作为目标，或者为了其自身的缘故而想要某事物，我们最有理由要做的事情就是任何能够最好地满足我们当下知情的目的式欲望的事情。由于斯加力和克里蒙森都有着充分的信息，且他们现在都偏好未来一个小时的剧痛而不

是一分钟的微痛，所以他们都认为最有理由选择具有这种痛苦。

我认为，基于这些假设，这些人的偏好和行动仍然是不合理的。斯加力和克里蒙森偏好一个小时的剧痛而不是一分钟的微痛，他们没有对明显的决定性理由做出回应。但他们的信念可能不是不合理的。虽然如下的信念是疯狂的：认为我们有理由关注未来的痛苦，但周二的痛苦除外；但这样的信念并非疯狂的：认为所有的实践理由都是由欲望给予的，并且我们没有理由为了事物自身之故而想要该事物。许多人接受这种主观理论是因为受到了这样的教导，他们的老师甚至提都不提任何客观理论。尽管我认为主观理论是错误的，但这些人接受这种理论可能并不是不理性的。

不仅如此，不像斯加力和克里蒙森，许多现实中的人具有合理的欲望和偏好。对于未来的福祉，尽管这些人认为他们没有任何理由要予以关注，但他们确实关注。他们可能对所有未来都是同样关注的，因此，如果他们认为推迟痛苦只会使得自己更为痛苦，那么他们永远也不会做这种推迟。这些人理性地回应了“给予他们理由要关注其未来福祉”的那些事实，他们确实以这种方式来回应这些理由。他们的错误仅仅在于，没有在自觉的层次相信他们具有这些理由。有些主观主义者甚至具有这种信念，并在学术之外的生活中依此行事，但是当他们教书或写作时就忽略或拒绝这些信念。（这就类似于许多经济学家：他们只是在写作和教书时，才相信福祉的人际间比较是没有意义的。）

123

18 有关合理性的其他观点

我们接下来简要地考察，我们有关欲望、目标和行动之合理性的其他观点。当某些人称某个行动“合理”的时候，他们意指这个行动最有可能满足我们当下的欲望，或更确切地说使我们的期望效用最大化。其他人则意指，这个行动对我们最有益，因此在一种更古老且时间中立的意义上使我们的期望效用最大化。我们可以宣称，这些行动在基于当下欲望和利己主义的意义上是“合理的”。如果人们在这些意义上使用“合理的”，那么他们确实能主张：若我们所做的事情确实会最大化我们的期望效用，最可能最有益于我们，我们的行动就是合理的。但这些主张并不是实质性的，因为实质性主张与有关合理性的其他观点是有可能相冲突的。而这些主张不过是告诉我们：我们以这样的方式行动时，我们是在这样行动。要提出实质性的主张，我们就必须在其他意义上使用“合理的”和“不合理的”。我一直主张，我们最好是在其日常意义上来使

用这些语词，以表达某类赞扬或批评。

大部分作者在提出有关合理性的实质性主张时，主要是讨论我们应当如何合理地努力满足我们的欲望，或实现我们的目标，而在很多情形中我们并不知道全部的相关事实。正如我说过的，这个问题具有很大的实践重要性，很多人对之做了很好的讨论。在这样的情形中我们如何行动才是合理的，如何才能最好地回应风险和不确定性，有些人对此提出了相冲突的主张。但这些分歧并不深刻。

对于何种欲望或目标才是合理的，讨论就更是少得多。当人们讨论这个更为根本的问题时，他们的分歧更为深刻。

基于一种普通的观点，若我们对某些欲望的持有会产生好的效果，则这些欲望是合理的。如果某个反覆无常的专制者威胁我，除非自现在起的一个小时内，我想要自己受折磨，否则就会折磨我，但这并不使得“我想要受折磨”这个欲望是合理的。这个专制者的威胁可能会使得“我尽力致使自己具有这个不合理的欲望”是合理的。那么，我具有这个欲望在某种意义上讲是合理的。但这个欲望本身仍然是不合理的。这将是一种合理的不合理情况。

正如我说过的，按照另一种普通的观点，当且因为我们的欲望或目标以正确的方式因果地依赖于合理的信念时，它们就是合理的。我已经论证过，我们要拒绝这种观点。

有些人主张，我们欲望的合理性在部分程度上依赖于有关欲望来源的其他某些事情。这些作者主张，如果这些欲望是通过自主慎思而形成的，它们就是合理的；它们如果是由于诸如被催眠或灌输之类的其他某些方式而形成的欲望，那就是不合理的。我认为，我们应当拒绝这样的观点。即使我们是因为被催眠或灌输而具有某些欲望，它们依然可以是合理的。例如，如果我们不怎么关注未来，那么我们可以通过催眠而具有这些合理的欲望。或者我们可以被灌输要爱我们的敌人，或想要至少日行一善。我认为，这样的爱和欲望是完全合理的。我们接下来假定，我们经由自主的慎思之后，想要饿死自己，从而失去一直过得很幸福的生活，或者具有其他的欲望，想要其他完全不可欲的事物。这些欲望的自主起源不会使得它们或我们成为理性的。相反，如果我们不是通过自主的慎思形成这些欲望，而是通过外部的干涉比如催眠而具有这些欲望，那么我们不理性的程度还没有那么的严重。

按照其他某种类似的观点，我们欲望的合理性依赖的不是我们如何获得它们，而是有什么因素可以致使我们失去它们，或者说它们是否经得起某些检验。

布兰德特表示，如果我们的欲望能够通得过某些认知过程或基于信念的心理治疗，那么它们是合理的。基于这些论说，如果我们疯狂得无可救药，我们的欲望就是合理的。这种主张没什么用处。

127 按照另外一组观点来说，当我们的欲望或偏好不一致时，它们是不合理的。若两个信念不可能同时为真，则它们是不一致的。由于欲望不可能为真，所以这个定义不能直接适用于欲望。但是许多作者写道，两个欲望可以是不一致的，其意思是我们不可能同时满足它们。

这种不一致性不包含任何不合理性。假定某个船舶失事后，我无法同时救两个孩子，只能救其中的一个。即使我认识到这个事实，我还是能继续理性地想要救我的两个孩子。如果我们知道两个欲望不可能同时得到满足，这可能会使得我们把满足两个欲望当作目标或意图是不合理的。但是这些欲望自身仍然是合理的，我们具有它们还是合理的。若我们的欲望在这种意义上是不一致的，这可能使得我们具有它们是不幸的。然而，正如我主张的，这并不使得这种欲望是不合理的。

不一致性要成为一种错误，就必须以其他方式定义。尽管欲望不可能有对错之分，但许多欲望依赖于有关好坏的信念；而这些信念可能是不一致的，因此不可能全部为真。当欲望依赖于这种不一致的规范信念时，我们就能主张，这些欲望在派生意义上是不一致的。

如果我们既想某事发生又想它不发生，情况看起来就是这个样子的。在具有这些欲望时，我们看起来是不一致地假定：这件事情发生既是更好的又是更坏的。但是在大部分这样的情形中，我们是在假定某个事件在某种意义上是好的，在另一种意义上是坏的。例如，我可能想要完成我一生的工作，这样避免去世时工作还没完成，同时我又不想完成我的工作，这样我在生前一直有重要的事可做。这样的欲望和规范信念没有任何不一致性。两个欲望在依赖信念的方式上要想是不合理地不一致的，它们就必须依赖于这样的信念：同一个事物以完全相同的方式既是好的又是坏的。我们是否有可能具有这样的信念和欲望，这是不清楚的。但如果有可能，那么诉诸不一致性的反驳在此就是有理的。

128 当我们转向更大组的偏好时，更可能出现不一致。我们可能偏好B多过A，C多过B，A多过C。这种偏好可以称作是循环的。如果这些只是单纯的偏好，不依赖于规范信念，那么我们是否可以称这些偏好是不合理的，这就是不清楚的。这个主张通常得到这样的辩解：我们如果具有这种循环的偏好就可能被人利用。例如我们可能愿意付三次钱，从而以B换A，以C换B，最后又以A

换C。我们的钱会被浪费，因为我们会回到了A这个起点。但是这个反驳诉诸的不是这组偏好的不一致性，而是其坏的效果。但我们若具有这样的偏好，也可能具有好的效果。假定，只要情境改变的方式是我所偏好的，这个改变就会给予我乐趣。如果我对于三种极易改变的情境X、Y和Z具有这种循环的偏好，这可能会对我有点好处。我们可以不停地进行这种循环，从每一次改变之中获取乐趣。这种快乐循环将成为某种快乐地运转的永动机。

当这种偏好依赖于某些规范信念时，情况就不同了。假定因为我们相信X内在地优于Y，Y优于Z，Z优于X，由此我们就具有那类偏好。“内在地优于”这种关系是传递性的，若这个主张言之成理（我则相信是真的），那么这组信念就是不一致的。基于这种观点，如果你比一个比我高的人更高，那么我就不能比你更高，同样，如果Y比Z好，Z比X好，那么X就不能比Y好。如果这组信念是不一致的，那么我们可以主张，这会使得这组偏好是派生地不合理的。尽管包含这种偏好的情形在理论上是有趣的，但我认为，它们在实践上的重要性并不大。

我一直主张，我们欲望的合理性既不依赖其来源，也不依赖于与其他欲望的一致性。持有相反观点的人，有些受到了信念与欲望之间具有可比性的误导。我们大部分信念的合理性确实要么依赖于来源，要么依赖于与其他信念的一致性，或两者都依赖。相对而言，只有极少信念的合理性只依赖于其内容：我们相信的事物。这只对那些必然为真或必然为假的信念（诸如数学或逻辑信念）而言，情况才是这样子的。有些信念是内在地不合理的，例如，我们相信的若是某些明显的矛盾。但是我们大部分的信念都是经验的和偶然的，意指它们是关于我们可观察的时空宇宙碰巧如何。有些经验信念的合理性只依赖于其内容。笛卡尔的信念“我存在”以及更慎重的佛教信念“此即心识”，或许是其中两例。这些信念也许是必然为真的，由此使得它们是内在地合理的。但是这类经验信念是极少的。有些经验信念（诸如某些精神病患者认为自己是拿破仑或维多利亚女王的信念）可能根据其内容看起来就是不合理的。但即使这些信念的不合理性依然大部分是由于其来源上的问题，或者是由于与其他信念不一致的问题。大部分经验信念的合理性不可能仅取决于其内容，因为仅当这种信念符合这个世界时，它们才为真。我们关于这个世界的信念，其合理性依赖于我们的其他信念、我们的感知经验以及其他可获得的证据。

上述主张都不适用于我们内在的目的式欲望。这些欲望的合理性并不依赖于它们是如何出现的，也不依赖于它们与其他欲望的一致性。当我们想把某

件事情作为目标，或者为了其自身的缘故时，这种欲望的合理性只取决于我们对此欲望对象的（或者我们想要之物的）信念。如基于价值的客观理论所主张的，这些欲望所依赖的信念若为真，就会使得其对象在某种意义上是好的或值得追求的，则这些欲望就是合理的。这才是核心且根本的道理，但我们一直在考察的大部分理论不是忽略它，就是否定它。

就拒绝在信念的合理性与欲望的合理性之间做出那些类比而言，我并没有忘记我们的许多欲望依赖于我们的规范信念。这些信念不是关于经验和偶然的真理，而是关于必然的真理。例如，不应得的痛苦本身就是坏的。这种规范信念要想合理，我们并不需要具有它们匹配于这个实际世界的证据，因为这些信念在所有可能世界都为真。

第六章 道德

19 西季威克的二元论

理由的各种客观理论存在若干差别。一个差别在于，它们所主张的“在蕴含理由的意义上有好坏之分的事件”的范围不同。有些理论主张，只有当相对于某个结果来说，另一个结果至少对一个人更差时，后一种结果才是更差的。我将论证，情况不是这个样子的。情况也并不是只有结果才是值得追求的，因为有些行动本身就是好的，而有些事情可能仅由于其自身之故就是值得做的。 130

我们有理由促进谁的福祉，各种客观理论对此的主张也存在着差别。我们接下来考虑理由的三种客观理论。按照

理性利己主义：我们总是最有理由去做对自己最好的任何事。

按照

理性的不偏不倚主义：我们总是最有理由去做任何不偏不倚地最好的事情。

如果从不偏不倚的视角看，我们所做的事情是每个人都最有理由想我们去做的事情，那么这样的行动就是在蕴含理由的意义上是不偏不倚地最好的。基于一种观点，不偏不倚地最好的事情就是从总体上讲对人们最好的事情，也就是最有益于人们的事情。

在其乏味的大作《伦理学方法》中，西季威克具体说明并综合了这两种观 131

点。按照西季威克所称的

实践理由的二元论：我们总是最有理由做不偏不倚地最好的事情，除非另有某种行动对自己是最好的。在此情况下，我们将有充分的理由按照其中任一方式行动。如果我们了解相关事实，那么其中任何一种行动都是合理的。

我认为在这三种观点之中，西季威克的观点是最接近真理的。按照理性利己主义者的观点，我们不可能有充分的理由去做对自己不利的东西，这是不对的。例如，当且因为某个行动可让事情变得在不偏不倚意义上好很多时，我们就有理由这样做。如果伤害自己是挽救他人生命的唯一方法，那么我就有充分的理由这样做。按照理性的不偏不倚主义者的观点，我们不可能有充分的理由去让事情在不偏不倚意义上变差，这也是不对的。例如，当且因为我们的行动能让我们自己好很多时，我们就有这样的理由。我有充分的理由挽救自己的生命而不是几个陌生人的生命。

按照西季威克的观点，我们的行动既具有不偏不倚的理由也有自利的理由，但这些理由是无法比较的。因此，一旦一种行动是不偏不倚地最好的而另一种是对自己最好的，那我们就有充分的理由以其中任一种方式行动，没有任何一种理由能被另一种理由所压倒。

有些理由是可以做精确比较的，意思是它们在相对的权重或强度上具有精确的事实。按照基于欲望的某些主观理论，一切理由都是可以精确地比较的，因为存在着关于我们一切欲望的相对强度之确切事实。按照基于价值的客观理论，当我们必须在非常类似的事物（诸如两个樱桃或两本同样的书）中进行选择时，我们具有完全一样的理由去选择（或许更好的说法是挑选）任何一样。

132 当我们比较某些种类的理由（诸如我们在买某物的成本差别所提供的理由，相同快乐或痛苦时间的长度差别所提供的理由）时，这些理由的强度就是能精确地可比较的。但当我们比较大部分理由时，无论是相同类型的还是不同类型的，这些理由就远不是能这么比较的。

下面两种情况所提供的理由就不是类似的：一种痛苦的强度更大，另一种痛苦的时间更长。如果我们必须在两种痛苦（一种是短暂的强烈痛苦，另一种则是长时但远没那么强烈的痛苦）之间做选择，两种体验之一可能更糟糕，意思就是我们更有理由偏好另一种。但我想说，在这些理由的相对强度上，可

能并不存在精确的事实。例如，其中的一种痛苦不可能比另一种痛苦要差2.36倍。即使在原则上，我们也并没有这样的一个标度（scale），可以根据它来精确地比较我们想避免这样两种不同痛苦的理由之强度。人们也许会挑战此类主张，因为痛苦的持续长度与强度是同一种坏处的构成成分。对于诸多其他不同类型的理由（如经济理由、审美理由、守诺理由或帮助陌生人的理由）的相对强度，也仅存在极不确切的事实。然而，这样的理由是可比较的，因为每类理由之中较弱的都可以比另一类理由中较强的更弱，或被其所压倒。

西季威克的二元论与此形成了对照，按照这种观点，不偏不倚的理由与自利的理由完全是不可比较的。没有任何不偏不倚的理由能够比任何自利理由要强或弱。这类观点极难捍卫。假定我们为了建造新房，要在某些建筑方案之间做出选择。当两个相冲突的理由没有一个比另一个强时，我们以两种方式行动都是合理的。如果经济理由与美学理由是完全不可比的，下面两者就将都是真的：

（1）我们能够合理地选择两个计划之一，因为它使得这个建筑要便宜一美金，尽管要丑得多。

（2）我们也能够合理地选择另两个计划之一，因为它使得这个建筑要略微好看一点，尽管要多花十亿美金。

我们也许能设想，这些选择如何能够是合理的，因为我们也许有理由可以给予这个建筑的美或成本以绝对优先性。但我们能够合理地做出这两种选择，这种主张是完全讲不通的。正如这个例子所显示的，我们想要捍卫西季威克的观点（不偏不倚的理由与自利理由是完全不可比的），只主张这些理由的类型不同是不够的。

西季威克对其观点的捍卫，在部分程度上诉诸了个人同一性的合理重要性。西季威克主张，鉴于每个人生活的统一性，在我们的整体生活中，我们每个人都有强有力的理由关注自身福祉。由于人与人的区分非常深刻，一个人的幸福损失无法为另一个人的幸福所得所补偿，这具有合理的重要性。西季威克在此诉诸了个人的分立性，这一直被人们看作是“伦理学的根本事实”。

西季威克的二元论也取决于内格尔所称的立场（standpoints）的二元性。我们是从自己的个人视角来过生活的。但是，我们也能仿佛拥有某种属于超然观察者的不偏不倚的视角，从而考虑到这个世界以及其中的一切人。西季威克主张，当我们询问我们最有理由做什么时，从这两种视角出发我们会得到不同

的答案。从我们自己的视角看，自利理由是最高的，意思是我们总是最有理由去做对自己最好的任何事。从不偏不倚的视角看，我们总是最有理由去做任何不偏不倚地最好的事情，由此不偏不倚的理由是最高的。

我们接下来假定一个行动是不偏不倚地最好的，另一个则对于我们自己是最好的。这里自利的理由与不偏不倚的理由是相冲突的。在这样的情形下，我们可以问：经过通盘考虑，我们最有理由做什么。但西季威克认为，这个问题永远也没有有益的答案。我们永远也不会更有理由做这两种行动之一。“实践合理性”会是自相矛盾的，无话可说，没法给我们任何指南。在西季威克看来，这个结论是让人极不满意的。

西季威克的推理似乎是这样的：

(A) 当我们试图决定我们最有理由做什么时，我们能够理性地从自己个人的视角或者是从虚构的不偏不倚的视角询问这个问题。

(B) 当我们从自己的个人视角问这个问题时，答案是自利理由是最高的。

(C) 当我们从不偏不倚的视角问这个问题时，答案是不偏不倚的理由是最高的。

(D) 当我们比较这两类理由的强度时，我们需要第三种中立的视角。

(E) 我们并没有这样的一种视角。

因此

不偏不倚的理由和自利理由是完全不可比的。当这些理由冲突时，其中任一类理由都不可能比另一类更强。

因此

在所有这些的情形下，我们具有充分的理由要么是做不偏不倚地最好的事情，要么是对做我们自己最好的事情。如果我们知道各种事实，两种行动就都是理性的。

我们能称此为两视角论证（two viewpoints argument）。

我认为西季威克的观点在部分程度上是对的。但我们应当拒绝这个论证，修正这种观点。

我们应该拒绝前提（A）。我们值得去追问，如果我们处于某个外在观察者的不偏不倚的立场，我们会最有理由想要或偏好什么。通过诉诸每个人都具有这种不偏不倚的理由想要或偏好什么，我们更易于解释：后果在一种重要的意义上是更好或更坏。但是当我们努力决定我们最有理由做什么时，我们应当从自己的实际视角来问这个问题。不能仅仅由于我们如果具有其他纯粹虚构的视角就不会有我们的实际理由，我们就忽略这些理由。

135

我们也应该拒绝（D）。为了能够对偏向性理由与不偏不倚的理由做出比较，我们并没有必要具有第三种中立视角。我们可以从实际的个人视角来比较这两类理由。

当比较这些理由时，我们接下来可以拒绝前提（B）。按照西季威克的观点，我们能够理性地做这样的事情：明知它对自己产生的好处很少，但对其他人产生的坏处要多得多。例如，我们能够理性地这么做：宁让自己免于一分钟的不适，也不救一百万人的生命或使他们免于极度痛苦。如果我们这样做了，其他人的主要反应会是恐怖和愤慨。但我们的问题是：这个行动是合理的吗？

有些人会做肯定回答。按照这些人的观点，如果我们知道这个行动会最好地满足我们的欲望或对我们是最好的，那么不管这个行动是多么恐怖，它都会是合理的。然而，持有这种观点的人之中，许多人要么是在“基于当前欲望”的意义上使用“合理的”，要么是在“利己主义”的意义上使用它。如果这些人主张某个行动是合理的，其中某些人的意思就是：在做最好地满足我们当下欲望的事情时，我们是在做能最好地满足这些欲望的事情；其他人的意思则是：在做对自己最好的事情时，我们是在做对自己最好的那些事情。这些无足轻重的主张完全可以被忽略。当我询问这个行动是否合理时，我不是在这些意义上使用“合理的”。我在询问这个行动是否应得一类批评。我认为，只有当我们具有“相关事实”的信念，且这些信念为真会给予我们充分的理由去做某件事情时，我们那样做才是理性的。

在我的虚构情形中，我们知道相关的事实。我们是否有充分的理由使自己免于一种轻度的不适，而不去挽救一百万人的生命或者是让他们免于极度的痛苦呢？我认为答案是否定的。这种恐怖的行动不可能是合理的。

我们可以补上：因为此类行动在道德上是不当的，所以它们不会是合理

136

的。西季威克假定，我们的自利理由不可能比我们避免行动不当的理由更弱或被其所压倒。我们应该拒绝这种假定。

我们也可能拒绝西季威克的这一主张：我们总是能理性地做已知会使事物变得最好的任何事情。西季威克作为一个行动后果论者，他认为这样的行动总是道德上正当的。我们大部分人会拒绝这种观点，因为我们认为，某些行动尽管会使事情进展最好但仍是不当的。我们可能主张，这种行动的不当性常常给予我们决定性的理由不这样做。

我很快会转向有关道德的问题，有关避免行为不当的理由问题。但我们可以首先用其他方式修正西季威克的观点。这种观点过分夸大了个人同一性的合理重要性。西季威克正确地主张，我们有理由特别关注自己的未来福祉。但我们也有其他的类似理由。在我们关注自己未来福祉的理由中，许多不是由“这个未来是自己的”这个事实所提供的，而是由“现在的我们”与“将来的我们”之间的各种心理关联所提供的。我们大部分人与某些他人（诸如近亲及爱人）在某种程度上都具有类似的关联。我会说，我们与这样的人具有紧密的纽带。我们与这些人的关联能给予我们理由，要特别关注他们的福祉。我们要有益于这些人的某些理由比关爱自己的某些理由要强得多。因此，我们应该拒绝西季威克的这一主张：当我们从个人视角来评价时，自利理由是最高的。

除了具有这些个人性理由和偏向性理由关注自己以及与我们具有紧密纽带的人的福祉之外，我们还有不偏不倚的理由关注每一个人的福祉。西季威克的某些主张似乎蕴含着：只有当我们从不偏不倚的视角考虑事情的时候，我们才有这样的理由。但情况不是这样的。内格尔设想自己是个利己主义者时写道：

假定我从火灾中得救，发现自己身处医院的烧伤病房。我想要止痛药，临床的那个人也是如此。他坦承希望我们都能得到吗啡，但我无法理解这一点。我理解为什么他有理由想自己得到吗啡，但他有什么理由想我也得到吗啡呢？是我的呻吟打扰他了吗？

正如内格尔的评论，这种利己主义的态度会“非常古怪”。除非我们受到教导而接受某种基于欲望的主观理论，或我们缺乏某种真正的“规范理由”之概念，否则，我们大部分人会正确地认为，我们有（不偏不倚的）理由想要减轻任何陌生人的痛苦。即使当我们的实际视角并不是不偏不倚的，我们还是具有这种不偏不倚的理由。正如我说过的，我们有益于陌生人的理由，不仅能够

与我们某些自利理由相冲突，还能比它们强得多。相对于使自己免于小伤害而言，我们挽救许多人的生命，使他们避免极度痛苦，这种理由要强得多。

然而，西季威克的观点在部分程度上是正确的。我认为，我们的偏向性理由和不偏不倚的理由仅仅可做非常不精确的比较。按照我们所称为的

广义的基于价值的客观观点：一个行动使得事情在不偏不倚的意义上更好，而另一个行动使得事情对自己或与我们具有紧密纽带的人更好，我们此时通常具有充分的理由做两个行动中的任何一个。

“通常”这个语词容许有各种各样的例外。对于什么时候我们没有充分的理由做两个行动中的任何一个，不同的广义的基于价值的客观观点可能会做出相冲突的主张。我认为，我们应当接受这样的一些观点。

为阐明这样的一种观点，我们能够假定，在

情形1中，我要么使自己不受某种伤害，要么挽救遥远之地的某个陌生人的生命。

以及在

138

情形2中，我要么救自己的命，要么救几个陌生人的命。

基于大部分人的观点，在两种情形中我在道德上可允许以两种方式之一行动。如果情况是如此，我认为，我在理性上也会允许以两种方式之一行动。在情形1中，我具有充分的理由使自己不受某种伤害，或者救陌生人的命。无论我的伤害是小到失去一个指头，大到失去双腿，我都有充分的理由。在情形2中，我有充分的理由救自己的命，或者救几个陌生人的命。无论陌生人的数目是2还是2000，我都有充分的理由。尽管我救两个陌生人的理由会比救2000个陌生人的理由要弱得多，但两种理由既不弱于也不强于我救自己命的理由。如果这样的主张是对的，这两类理由的相对强度将是非常不精确的。

我们可以主张，之所以会出现这么大的不精确性，是因为这些理由是由极为不同类型的事实所提供的。不偏不倚的理由是行动者中立的（agent-neutral），意思是这些理由是由“其描述并不需要参考我们”的事实所提供

的。例子之一是某个事件会导致极大的苦难这个事实。我们有理由对任何人的痛苦表示遗憾，且我们有理由尽力阻止或减轻这种痛苦，而不管受苦者是谁、与我们关系如何。我们有理由防止任何有感觉者或有意识者的痛苦，并对这种痛苦表示遗憾。当我们处于痛苦时，正如内格尔所言：

在思想中，痛苦可以与它是我的痛苦这个事实相分离而丝毫不减其可怕性……苦难是件坏事，无需多言；不仅仅对于受苦者是如此……这种体验不应当继续下去，不管谁是受苦者。

139 相对比而言，我们的个人性理由和偏向性理由是行动者相对的（agent-relative）。这些理由是由“其描述必定参考我们”的那种事实所提供的。我们每个人都有理由要特别关注自己以及与我们具有某种关联的人的福祉。尽管我有理由既防止自己也防止任何陌生人的痛苦，但我与自己的关系，与自己痛苦的关系，完全不同于我与陌生人的关系，与陌生人的痛苦的关系。这就是为什么只能对这些理由做如此不精确的比较。

按照某种广义的基于价值的观点，当我们在道德上可允许的行动之间进行选择时，我们给予自己某种利益的理由，总是强于或压倒了我们给予陌生人同等利益的理由；但这个差别是非常不精确的。基于某种这样的观点，理性要求我们给予自己福祉的权重要比给予陌生人福祉的权重大，但是这种权重比可以小到2，也可以大到数百，甚至数千。

我认为，这些观点太过于利己主义了。我们也通常能给予陌生人的福祉以同样的权重，甚至是更大的权重。假定我像内格尔一样身处某个病房，唯一剂量的吗啡是属于我的。我认为，我具有充分的理由把这个吗啡给予临床的那个人。即使这个陌生人的疼痛没有我的厉害，我还是有充分的理由这样做。

人们可能会主张，仅当自己放弃的利益相当小，这样的行动才是合理的。相反假定，在

船舶事故1中，我有一个救生筏，既可救自己的命，也可救某个陌生人的命。这个陌生人与我在相关方面是差不多的，这样我们的死亡对各自的损失是差不多的。

当筹码是这样大时，理性似乎要求我们给予自己福祉的优先性或权重要

大得多。如果情况如此，那么我就没有充分的理由，去救这个陌生人而不是自己。这个行动即使是道德上可敬的，但不会是完全合理的。

我倾向于认为这个行动可以是完全合理的。这个陌生人的福祉与我的福祉是同等重要的。如果我放弃自己的命来救这个陌生人，那么这将是慷慨的善举。我认为，这些事实给予我充分的理由这样做。

我必须承认，这种观点会受到强有力的反驳。我认为正如西季威克所主张的，我们有理由特别关注自己的福祉。在这个虚构例子中，我的死亡与陌生人的死亡在不偏不倚的意义上是一样坏的。既然我有同等的理由来救自己或是陌生人，那么我就可以主张，我的自利理由会打破这个僵局或扭转局势，从而综合考虑，这给予我决定性的理由救自己。

然而，这些理由可能不是决定性的。即使筹码非常大，理性可能并不要求我们给予自己福祉以任何优先性。我们可有证成地接受西季威克观点的某种修正版本。按照我们所称的

140

广义二元论：若我们在道德上可许可的两个行动之间选择，其中一个对我们自己更好，另一个则对一个或更多的陌生人更好，则我们能理性地给予自己福祉以更大的权重，或者给予每个人的福祉以大致相等的权重。

这种观点的不同版本则会提出进一步的不同主张。尽管这样的观点并不在理性上要求我们给予自己的福祉以更大的权重，但它允许我们给予自己的福祉以大得多的权重，并且确实要求我们不给予陌生人的福祉以更大的权重。例如，基于这种观点的某些版本，我们救自己的手指头而不是陌生人的命是合理的，但救陌生人的指头而不是自己的命是不合理的。这种广义二元论允许我们给予自己的福祉以更大的优先性，但是要求我们不给予陌生人的福祉以这种优先性，它认可并且支持我们有理由要特别关注自己的福祉。

我们接下来假定，在

船舶事故2中，我可以救某个陌生人的命，或者救与我具有紧密纽带的人（如孩子或朋友）的命。

正如广义二元论者的主张，我不能理性地选择救这个陌生人。我在道德上

141

应当给予我的孩子以优先性。我这样做会具有其他强大的非道德理由，诸如我对孩子或朋友的爱所提供的理由。如果我救了这个陌生人而不是我的孩子或朋友，那么这不会是慷慨的善举。

这也同样适用于船舶事故¹的情形。我也许会有年幼的孩子，他们要靠我，或者对其他人具有某种道德义务。这可能会使得我救陌生人而不是自己是不当的，因为我无法关爱我的孩子或履行我的义务。这个陌生人可能具有类似的义务，他的死亡也会使得它们无法被履行，但这些义务不是我的义务。如果我的死亡对我爱的人和爱我的人是坏的，这会给予我其他决定性的理由要救自己的命。这样一来，在船舶事故¹情形的这个版本中，理性会要求我救自己。

我们接下来假定，我与其他人没有任何这样的纽带，可给予我理由或产生这种义务。我倾向于认为，在情形的这个版本中，我能够理性地选择放弃自己而救陌生人。在这样的情形中，理性可以允许我们忽略要特别关注自己福祉的理由。但我们没有必要决定情况是否如此，即尽管我的行动是道德上可敬的，但它不是完全合理的。

20 最深刻的问题

我们现在可以转向理由与道德之间的关系。按照

道德理性主义：我们总是最有理由履行我们的义务。我们以任何自认不当的方式行动都是不合理的。

按照

理性利己主义：我们总是最有理由做对自己最好的事情。我们以任何自认会不利于自己的方式行动都是不理性的。

142 这两种理论是许多人都接受的。许多人相信，我们每个人都有来世，无论是否履行义务，我们都会得到应得的幸福或苦痛。因此，义务和自我利益是从不冲突的。世上大多数伟大的宗教都持如此主张。

西季威克怀疑我们具有这样的来世；并且他认为，在某些情形中，义务与自我利益很可能是冲突的。西季威克主张，如果存在着这样的情形，这就会引

起“伦理学中最深刻的问题”。

西季威克的问题在部分程度上在于：在他看来，不管是道德理性主义还是理性利己主义，它们在直觉上都是说得通的；但如果义务与自我利益有时候是冲突的，这些观点就不可能都为真。如果我们必须在两个行动之间做选择，其中一个是我们的义务，另一个则更有利于自己，这些观点就蕴含着我们最有理由以其中任一种方式来行动。这是无法设想的或者说逻辑上不可能的。正如我们不可能在两个钱包中都放大多数的钱一样，我们也不可能最有理由以两种不同的方式去行动。这样一来，若义务和自我利益有时候是会冲突的，我们就不得不放弃或至少修正一种观点。

有的作者在考虑这两个选项时会拒绝道德理性主义。例如，里德（Reid）主张，如果履行义务不利于自己的利益，我们就会“归结为这样一种可悲的两难选择：最好的情况不是做无赖就是做傻瓜”。我们如果不履行义务就是无赖，如果履行就是傻瓜。其他的作者则拒绝理性利己主义。按照这些人的看法，我们永远都没有充分的理由做不当的行为，即使这是唯一的方式让自己免于严重的痛苦甚至是死亡。

西季威克发现这样的主张是令人难以置信的。他不是拒绝其中之一，相反是修正两者。按照西季威克的另一版本的二元论，我们可以称为

义务与自我利益的二元论：如果义务和自我利益永不冲突，那么我们总是最有理由既履行我们的义务也做最有益于自己的事情。但如果我们不得不在两个行动之间选择，一个行动是我们的义务，但另一个行动对自己更好，理性就无法提供任何指南。在这样的情形下，我们就不会有更强的理由以其中任一方式行动。如果我们了解相关事实，两个行动就都是理性的。

143

部分原因是接受了这种观点，西季威克热望义务与自我利益永远不冲突。他写道，

如果存在着这种冲突，我们对于行动之内在合情理性的整个信念体系就会坍塌……由此义务的秩序实际上归于混沌，人们为了构架理性行动的完美理想而做出的长期的智识努力，看来命中注定会失败。

然而，这些令人印象深刻的严峻主张是言过其实的。西季威克认为，义务与自我利益在大部分情形下并不冲突。西季威克的观点蕴含着，在许多这样的情形下，我们会最有理由履行我们的义务，且不用付出任何代价。在这样的世界之中，义务的秩序不会变为混沌。我们若得出，义务和自我利益相冲突时我们就能合情理地或合理地以其中任一种方式行动，则我们有关何为合情理行动的整体信念体系也不会坍塌。但若在此类情形中，我们和其他人有充分的理由做出不当的行动，那就是糟糕的。我们可能会说，道德主义者的问题在于我们是否能够避免这个结论。理性若在这样的情形中不能给予任何指南，则会让人很失望。我们可能指望，至少在某些这样的情形下，有些事情是我们最有理由要做的。我们也许会说，理性主义者的问题就是：这是否为真。

这些问题还可能以其他的形式出现。西季威克假定，如果我们具有充分的理由做不当之举，这些理由就会是自利性的。由于我们还能具有其他强有力理由做不当之举，我们就不应该做出这样的假设。有些这样的理由是个人性和偏向性的，但不是自利性的。如果某个不当之举不是救自己，而是救我们的近亲或所爱之人，或使他们免于巨痛，我们就有充分的理由做不当之举。

144 我们可能也有强有力的不偏不倚的理由做出不当之举。作为一个行动后果论者，西季威克主张我们总是应当做出使事情变得最好的任何行动。我们大多数人都拒绝这种观点，因为我们认为，某些行动即使会使事情进展最好也是不当的。例如，即使杀死一个人是挽救许多其他人的唯一方式，这可能依然是不当的。然而，即使这样做是不当的，但人们仍然可以主张，这个事实——我们在救众多他人的生命，由此使事情进展最好——给予了我们充分的理由这么做。如果情况真是这样，这就是我们可以行动不当的另一种情形。

还有第三种可能性。基于西季威克的观点，我们总是具有充分的理由履行我们的义务和避免行动不当。我们可称此观点为弱道德理性主义。如果我们是理由的主观主义者，我们必定会拒绝这种观点。例如罗尔斯就主张，如果我们当下知情的欲望可以通过不义之举而得到最好的满足，我们就没有充分的理由去行正义之举。按照此类主观理论，我们不仅可能没有任何充分的理由履行义务，甚至可能还有决定性的理由行不当之举。由此，我们履行义务就可能是不合理的。

为了涵盖这些不同的可能性，我们可以修正西季威克对他所称“最深刻问题”的描述。当我们在不同的行动之间选择时，我们问：

Q1：我最有理由做什么？我有充分的或决定性的理由以任何这

样的方式行动吗？

Q2: 我在道德上应当做什么？有任何这样的行动是不当的吗？

这些问题似乎具有相冲突的答案，因为我们有时候具有充分的或决定性的理由做不当之事。我们的问题是要决定我们是否确实具有或能够具有这样的理由，如果这为真，那么我们应得出何种进一步的结论。

在考虑这样的问题时，区分两种观念下的规范性是有益的。按照涉及理由的观念，规范性涉及理由或显见理由。按照涉及规则的观念，规范性涉及规则或要求，它们区分什么是正确的或不正确的、允许的或不允许的。例如，有些行动是法律、荣誉规则、礼节或语言规则所要求的。不纳税是违法的，不还赌债是不光彩的，以勺子吃豌豆、只写一个t拼写“committee”和用“驳斥”意指“否认”等是不正确的。此类要求或规则有时被称为“规范（norms）”。

145

这两种观念下的规范性是极为不同的。按照涉及规则的观念，我们引入或使人们接受某个规则，从而能创造出新的规范性事实。立法者能够立法，任何人都能造出定义新游戏的规则。莎士比亚写作的那个时代，英语单词只存在拼写规律而不是拼写规则，后来的英文作者才造出了这样的规则。相对比而言，基于涉及理由的观念，只有存在规范性理由或显见理由时，规范性才存在。我们无法通过让人们接受某条规则来造出这样的理由。

这两种观念可能会冲突。这样的规则或要求存在时，我们可能有理由要遵循它们，但这些理由大多不是单纯由这些规则的存在或接受所提供的，而是由其他的事实所提供的，其中大部分事实依赖于人们对这些规则的接受。如果我们在按章开车，撞车的概率就比较低。如果我们正确地拼写与使用语词，这会让我们显得更有教养，也更易理解。若没有给出理由的这些事实，我们可能就没有任何理由遵循这些规则或要求。例如，我们可能没有任何理由遵循某种时尚，或者不去违背某种禁忌。年幼的我被告知禁止做某事而问为什么时，若答案只是不要做这种事情，我就极为愤怒，因为这种回答并没有给出任何理由。

不过，许多这样的主张并不适用于道德要求。基于某些观点，正是我们创造了这些要求。我认为，这种观点仅在有限且通常是肤浅的意义上才是正确的。我们所能创造的只不过是特定的形式，是更根本、更普遍且非创造的要求在不同的共同体中所采取的形式。例如，人们应当照顾诸如小孩和由于疾病或年老而丧失能力者的这些人——各处的情况都是如此。尽管在大多数共同体中，正是由最亲的人承担这样的责任，但并不是所有地方都这么做。

还存在着各种各样非被创造的理性要求。例如，若我们相信有决定性的理由要做某个行为，理性就要求我们要么这样做，要么放弃这个信念。如果我们相信某个行动是实现某个目标的唯一手段，理性就要求我们要么这样做，要么放弃该目标。

道德要求常常与其他类型的要求相冲突。例如，法律可能要求我们行道德上的不当之举。许多男人相信，尽管决斗是道德上不当的，但不决斗却是不光彩的。我们大多数人认为，在这两类要求中，道德要求是更重要的。这些要求通常被称为是压倒性的。但这样的主张——道德要求在道德上是更重要的，或道德上压倒性的——是无足轻重的。法律要求在法律上是压倒性的，荣誉规则就这种规则而论是压倒性的。对于这些相互冲突的要求之相对重要性，我们要有能力提出有意义的主张，就需要某种不偏不倚的中立判准。

理由提供了这样一种判准。我们尊重这些要求的理由，其强度是我们能比较的。进行决斗的人，最多也只有很弱的理由遵循荣誉规则，但他们有很强的道德理由不去决斗。当我们受法律要求去行不当之举时，我们可能有决定性的道德理由违背法律。由此，道德要求在蕴含理由的意义上比荣誉规则或法律的要求更为重要。

理性要求在理性上是压倒性的，这种主张同样也是无足轻重的。这样我们应该询问，我们是否有理由遵循这些要求。这些要求在蕴含理由的意义有多重要，这是个难题。遵循这些要求可能是好的，但不是其本身就是好的，而仅仅是作为手段是好的。在诉诸关于“在蕴含理由的意义上何者重要”的主张时，我们并没有假定理性是重要的。

我们接下来要提到在道德要求与理性要求之间的一个差别。当决定要做什么时，我们常常该问是否有任何行动是道德上所要求的或不当的，但没必要问哪个行动是合理的。后一问题在下述情况下才会出现：我们考察自己过去的行动或他人的行动，且当我们询问这些行动是否使自己或他人要受到某类批评。相对于我们应当做什么或有理由做什么的问题而言，有关合理性的问题远没有那么重要。

正如我说过的，当我们决定要做什么时，我们有两个主要问题：

Q1：我最有理由做什么？

Q2：我在道德上应当做什么？

这些问题中，有关理由的问题是更宽泛也是更根本的。如果我们通常有决定性的理由行不当之举，这些问题的答案就通常会相冲突，而这就会破坏道德。道德要是重要的，我们就必须有理由关注道德且避免行动不当。任何这样

的主张反过来都不适用。如果我们有决定性的理由行动不当，那么这些行动的不当性就不会破坏这些理由。

人们可能会否认这样的主张。当我主张这些行动的不当性并不会破坏这些理由时，我的意思是我们仍然具有这些理由。人们可能会做出类似的主张，即使我们有决定性的理由行动不当，道德也不会受到破坏，因为这些行动仍然是不当的。

对道德的这种捍卫是很软弱的。我们能够类似地主张，即使我们没有任何理由遵循荣誉规则或礼节规则，这种规则也不会受到破坏。不决斗仍然是不荣誉的，用勺子吃豌豆仍然是不正确的。这些主张尽管是真的，却是无足轻重的。如果没有任何理由去做荣誉规则或礼节所要求的行动，这些要求就没有任何重要性。如果没有任何理由关注道德，或避免行动不当，道德就同样没有任何重要性。道德就可能以这样的方式遭到破坏。

人们接下来可能会反驳，我在提出这些主张时诉诸了涉及理由的重要性准则。我假定了仅当且因为我们或其他人有理由关注某个事物时，这个事物才是重要的。但我还没有捍卫这个准则。就像道德或荣誉规则一样，涉及理由的准则也不能自我支持。“道德在道德上是重要的，理性在理性上是重要的”，这样的主张是无足轻重的；同样，“理由在蕴含理由的意义上是重要的”，这样的主张也是无足轻重的。

148

正如这个反驳所正确地主张的，我们不可能通过诉诸有关理由的主张来表明理由是重要的。但是证成总要有止步的地方。如果理由是根本的，我们就不该指望，我们能够诉诸某种更深刻的其他判准来证成涉及理由的重要性准则。

我相信，理由是根本的。仅当我们或其他人有理由关注某个事物时，这个事物才是重要的。如果道德不是在这个意义上重要，就会有极大的影响，因为我们由此就没有任何理由关注行动是否得当。我们为了解释或捍卫道德的重要性，努力表明或主张我们确实具有这样的理由。既然我们总是有决定性的理由履行我们的义务或避免行为不当，道德也许就在蕴含理由的意义上具有最高的重要性。但是如果以这种方式捍卫道德的重要性，我们就必须承认，最深刻的问题不是我们在道德上应当做什么，而是我们有充分的或决定性的理由做什么。

本卷的余下部分将主要探讨道德。如果正如我刚说过的，理由是更根本的，看起来我就应该继续探讨理由。但我们有充分的理由转向道德。

第一，我们能言之成理地假定，我们确实具有强有力的理由关注道德和避免行为不当。在讨论道德时，我将在部分程度上探讨这些理由。这些是最有必要探讨的理由，它们引起了最为困难的一些问题。

第二，要对避免行动不当的理由所具有的强度能够做出判断，我们先必须回答关于何种行动是不当的某些问题。例子之一是这个问题：是否如行动后果论者所认为的，我们如果能够舍己而救几个陌生人的生命，那就应当牺牲自己的生命。如果这是真的，我们就能更有道理地主张，我们可能具有充分乃至决定性的理由行为不当。

按照大多数人所接受的信念集的重合，亦即西季威克所谓常识道德，我们在道德上允许极为重视自己的福祉。我们可能没有任何义务牺牲自己的生命，即使我们能因此救许多陌生人的生命。如果道德的要求在这种方式上远没有那么苛刻，那么我们主张能具有充分乃至决定性的理由行为不当，这就不那么说得通。

当我们考虑道德时，我们还有其他方式来考虑理由。基于几种可行的道德原则或理论，某个行动是否不当取决于在某种实际或虚构的情境下，我们或其他人最有理由或充分的理由同意、赞同、想要、选择或做什么。我们为了知道这些原则或理论蕴含什么，就必须回答有关理由的问题。这就像我们为了知道原子的本质与属性，我们就必须回答亚原子粒子的问题一样。

第七章 道德概念

21 无知或错误信念下的行动

我们探讨“何种行动是不当的”之前，下述做法是有益的：探讨“不当的”所指为何，即当我们相信某个行动不当的时候，我们在相信什么。这些问题是关于“不当的”的道德意义，以及这些意义所表达的各种概念。我们可以忽略其非道德的意义，诸如我们可能对某问题给出错误的答案，或者说在错误（wrong）一端打开麦片粥包。 150

人们常常假定“不当的”语词只有一种道德意义。我们考虑的行动者若是知道所有相关的道德事实，则这个假设是最可行的。我们起始就可假定，当我们思考这样的行动时，我们是在相同的意义上，也就是我们所称的日常意义上使用“不当的”。然而，我们在许多情形中并不知道所有相关的事实，必须在无知或错误的信念之下行动。当我们思考此类情形时，“不当的”的使用有几种（部分程度上）不同的意义。其中一些意义可以通过日常意义来界定。我们有些行动会是

在事实相对的（fact-relative）意义上是不当的，正好当我们若知道所有道德上相关事实，则这个行动会在日常意义上是不当的；

在信念相对的（belief-relative）意义上是不当的，正好当我们有关这些事实的信念若都为真，则这个行动在日常意义上是不当的；

以及

在证据相对的 (evidence-relative) 意义上是不当的, 正好当我们若相信的是已有证据给予我们决定性理由要相信的事情, 且这些信念为真, 则这个行动在日常意义上是不当的。

行动若在这些意义上都不是不当的, 就是正当的, 或至少是道德上许可的。若这些行动之外的所有选项在这些意义上都是不当的, 则它们就是我们在道德上应当做的事情。

有些作者主张或假定, 即使我们所考虑的行动者并不知道所有道德上的相关事实, 只要询问这些人的哪些行动在日常意义上会是或曾经是不当的, 这就足够了。其他的作者则假定, 我刚刚定义的诸种意义之一就是日常意义。我认为, 这些假定是错误的。我们应当在所有这些意义上使用“不当的”。我们如果不做出这些区分, 或只使用某一种意义, 那么就不会认识到某些重要的道理, 与其他人就会产生不必要的分歧。

为了阐明这样的观点, 我们假定: 作为你的医生, 我必须在不同的治疗方案之间进行选择。我是一个庸医, 因为现有证据表明, 对于不同的治疗方案具有什么样的可能效果, 我持有的诸多信念都是靠不住的。我还具有某些理由希望你死掉。这个故事可以用几种方式来继续。假定, 在

情形1中, 我给出了某个治疗方案, 我相信并希望它会救你的命, 但如其几乎肯定会产生的效果, 它事实上要了你的命,

以及, 在

情形2中, 我给出了某个治疗方案, 我相信并希望它会杀死你, 但如其几乎肯定会产生的效果, 它事实上救了你的命。

按照某些人的观点, 我们在信念相对的意义上使用“正当的”与“不当的”就够了。基于这种观点, 我们做出下述主张就足够了: 我在情形1中的行动是正当的, 因为我做的是我相信会救你命的事情; 我在情形2中的行动是不当的, 因为我做的是我相信会杀死你的事情。

只做出这样的主张是不够的。我们还应主张, 在情形1中, 我在事实相对与证据相对的意义上行动不当, 因为我的方案杀死了你, 正如基于现有证据它

几乎肯定会这样。如果我就应当怎么做而咨询某个有足够见识的人，那么这个人肯定不会说要用“他或她明知肯定会杀死你”的这个医疗方案。在情形2中，我应该类似地主张，我在事实相对与证据相对的意义上行动得当，因为我的行动救了你的命，正如其几乎肯定会产生这样的效果。我所做的正是任何有见识者会告诉我应当去做的事情。

接下来假定，尽管某种治疗几乎总是治好具有你那种特殊疾病的人，某种其他的治疗几乎总是会杀死这样的人，但你的情形是一种不可预测的例外情形。我们还假定，在

情形3中，我给你的治疗几乎肯定会杀死你，但它救了你的命——这个结果是我所希望的，也是我无根据地相信会如此的。

并且在

情形4中，我给你的治疗是几乎肯定会救你的命，但它杀了你——这个结果是我所希望的，也是我无根据地相信会如此的。

按照某些人的观点，我们在证据相对的意义上使用“正当的”与“不当的”就够了。基于这种观点，如果巫术信仰者试图通过用图钉钉入蜡人来杀死敌人，这个人不会行动不当。既然没有任何证据表明把图钉钉入蜡人有任何危害，因此这样做并无不当之处。我在情形4中给予你的治疗，基于现有证据几乎肯定会救你的命，故我的行动是正当的；但我在情形3中给予你的治疗，基于现有证据几乎肯定会杀死你，故我的行动是不当的。

正如前面一样，我们只这样主张是不够的。我们在情形4中不应该只说我行动得当，既然我的行动几乎肯定会救你的命；我们还应当主张，我是在谋杀你，因此在信念相对和事实相对的意义上是行动不当的。我们至少应该提到谋杀犯。

153

我们在情形3中只这样说也是不够的：我所做之事几乎肯定会杀死你，因而我行动不当。我们还应当主张，既然我想救你的命，我在事实相对和信念相对的意义上就是行动得当的。就我不认为我的行动几乎肯定会杀死你而言，我可能会犯有过失罪，因为我没有阅读最近的医学杂志，而这是我本该做的。但是反过来也可能是真的，我可能认真阅读了这些杂志，我的错误只在于，没有

相信这些杂志所报告的证据给予我决定性理由去相信的事情。尽管我医疗上的无能应该受到指责，但我没有对这些认知理由做出回应并非道德上的不当。

按照其他某些人的看法，我们在事实相对的意义使用“正当的”与“不当的”就够了。但假定，在

情形5中，我有证成地相信，我给你的治疗几乎肯定会救你的命，但它事实上杀了你。

我杀了你，所以我行动不当，但只这样说是不够的。我们还应当主张，我在信念相对和证据相对的意义行动上行动得当。我有证成地相信，我的行动几乎肯定会救你的命，这在道德上是重要的。反过来假定，在

情形6中，我有证成地相信，我给你的治疗几乎肯定会杀死你，但它事实上救了你。

我救了你，所以我行动得当，但只这样说也是不够的。我们还应主张，我在信念相对的意义是行动不当的，因为我认为我的行动会杀了你，这正是我想它做的。我们至少应该提到未遂的谋杀犯。

在不使用这些不同意义的“正当的”与“不当的”的情况下，我们也可以做出这些区分。我们可以只使用证据相对的意义，那么可以主张，尽管在情形4中我在谋杀你时没有行动不当，但我在道德上有决定性理由不这样做；而我的行动是可责备的，这会给予我悔恨的理由和给予他人愤慨的理由。或者我们可以只使用信念相对的意义，那么可以主张，尽管在情形1中我试图救你没有行动不当，但我在道德上有决定性的理由不这样做，因为我的行动杀了你，而我本该知道这样做肯定会杀了你。或者我们只使用事实相对的意义，于是可以主张，尽管在情形6中我救了你并非行动不当，但我的行动是可责备的，因为我试图杀了你。但是如果我们只在三种意义之一上使用“不当的”，那么在其他两种意义上使用“不当的”的那些人就可能会误解我们。我们和其他人可能会错误地认为彼此之间有分歧。我们考虑的情形之中若人们并不知道道德上所有相关事实，则并不存在任何一种意义的“不当的”，它是每个人都使用的。由此，我们最好是区分并且使用这三种意义。

我们接下来问何种意义是最重要的。正如我的主张所蕴含的，这取决于

我们询问的是哪个问题。我们可以从可责备性问题开始，我们可以将其看作是包含有关悔恨理由和愤慨理由的问题。这里最重要的是：人们在行动时相信什么。我们应该主张：

(A) 某个行动在信念相对的意义上是正当的，因为若该行动者的非道德信念为真，则该行动正当；在此情况下，这个事实就使该行动成为可责备的。

例如，在情形2、4和6中，我是按照我相信会杀死你的方式行动。如果我的信念为真，那么这些行动就都是正当的，因为蓄意谋杀是正当的，这样(A)就正确地蕴含着这些行动都是可责备的。

人们可能主张：

(B) 某个行动在事实相对意义上是正当的，因为若该行为者知道相关的事实，则这个行动就是正当的，那么在此情况下，这个事实就使得该行动是可责备的。

但是我们应当拒绝这个主张。回顾在

情形5中，我有证成地相信我做的事情会救你命，但它杀了你。

如果我知道这个行动会杀死你，这个行动就是正当的——既然如此，(B)就蕴含着这个行动是可责备的。但这显然是错误的。我若知道杀了你，会感到惊骇。但由于我有证成地相信我的行动会救你的命，故这个行动不是可责备的。我没有任何悔恨的理由，其他人也没有愤慨的理由。

这里对(B)有一个更广泛的反驳。假定在

情形7中，我有证成地相信我做的事情会救你的命，事实上它也救了你。

在这个情形中，我的行动非常清楚地不是可责备的，因为这个行动在任何意义上都不是正当的。尽管我的行动在情形5中杀了你，但在情形7中救了你；

从我的视角看，这个差别完全是运气问题。我从自己的视角称这个差别是运气问题，我意指的是，我此前不可能知道这个行动会杀死你，且我绝对无法控制这个事实。尽管这两个情形之间的这个差别完全是运气问题，但（B）蕴含着：我的行动在情形5中是可责备的，而在情形7中不是可责备的。（B）因此蕴含着

（C）行动的可责备性可能完全取决于运气。

当孩子们学习什么使得行动是可责备时，有些孩子具有的信念假定或蕴含了（C）。例如，有些孩子相信，若良好意愿的行动具有坏的效果，则即使这些效果是完全不可预测的，它们也是可责备的。有些成年人也具有类似的信念，诸如我们能够继承这种可责备性，为我们先辈的罪孽而内疚。这些罪孽是我们无法控制的。但是当我们更好地理解了可责备性时，我们意识到（C）是错误的。既然（B）蕴含了（C），我们就应当拒绝（B）。某个行动在事实相对的意义上是不当的，这一点并不使得这个行动是可责备的。

（C）存在着两个替代项，按照我们称为的

康德式观点，行动的可责备性不能取决于运气。

156 按照

准康德式观点，行动的可责备性不能完全取决于运气。但是当两个行的可责备性并不取决于运气时，那么其中某个行动可能确实因运气而更可责备。

我们很难明白，可责备性如何可以在部分程度上取决于运气，所以这个观点比康德式观点更讲不通。但是人们有时候又主张，这种准康德式观点具有的含义更加可行。我们返回到

情形2，其中我认为自己的行动会杀死你，实际上却救了你。

以及

情形4，其中我认为自己的行为会杀死你，也确实杀了你。

这两个行为在信念相对的意义都是不当的，因为如果我的信念是对的，那么这些行为就会正如我所希望的那样杀死你。然而，在事实相对的意义，我的行动只在情形4中是不当的。尽管我的行为只在情形4中杀死你，而在情形2中却救了你，但从我的视角看，这个差别完全是个运气问题。这样一来，这些行动按照康德式观点是同等可责备的。按照某些准康德主义者的观点，情况并不是这样的，他们认为

(D) 若行为在信念相对的意义不当而可责备，且这些行动也在事实相对的意义不当，则它们更可责备。

按照这种观点，我想杀死你的两个企图都是可责备的，但我的行动在情形4中更可责备，因为这个企图成功了。尽管谋杀未遂是可责备的，但成功的谋杀是更值得责备的，我和其他人有理由更加悔恨和愤慨。

有些准康德主义者还主张

(E) 行为若因为在信念相对的意义是不当的而可责备，且它们也在证据相对的意义不当，则更可责备。

157

但我们回顾，在

情形4中，我的治疗杀死了你，而基于现有证据，它几乎肯定会救活你，但我没有证成地相信它会杀死你。

我们接下来假定，在

情形8中，我的治疗杀死了你，且我有证成地相信它会杀死你。

这两个行动在信念相对的意义和事实相对的意义都是不当的，因为它们都杀死了你，并且我也相信它们会杀死你。(E) 蕴含着在情形8中，我的行动更

可责备，因为这个行动也在证据相对的意义上是正当的。我认为，我们应当拒绝这种主张。我们认为谋杀比谋杀未遂更可责备，这是讲得通的。但是如果且因为一种谋杀的行为者对于行为的可能后果有更好的信念，因为这些信念能得到已有证据更好的支持，我们因此认为它比另一种谋杀更可责备，这是讲不通的。我们最多只能是主张，如果潜在的谋杀者具有这种更有证成的信念，那么他们就更危险，因为他们的杀人企图更有可能成功。但这并不是在可责备性上的差别。

按照康德式观点，所有这种谋杀企图都是同等可责备的，不管其行为是否成功，或者成功的可能性如何。下述做法的可责备性是一样的：枪击某人且击中、枪击某人但没中、用针扎纸人且没证成地相信这会成功地杀死人。我们不会认为：仅仅因为我们在击中意向对象上不是那么成功或者认知上不那么合理，从而应得的责备就较少。

我认为，康德式观点是对的。尽管我们可以可行地认为，谋杀比谋杀未遂更值得责备，但我认为，这种主张的可行性可以通过其他方式得到充分的解释，我在注释中提到了一些这样的方式。

158

我们接下来可以定义第四种不当的方式。某个行动是

在道德信念相对的意义上是正当的，正好当行为者认为这个行动在日常意义上是正当的时候。

按照一种相当可行的观点，我们可以称为

托马斯主义观点：若人们认为自己的行动是正当的，这就足以使得其行动是正当的，即使这个行动若非如此就不是正当的。

例如，假定人工避孕、早期流产或助人安乐死不是正当的。基于托马斯主义者的观点，如果这样做的人们错误地认为这样的行动是正当的，那么这些行动就是正当的。然而，正如托马斯主义者所补充的，当人们认为某个行动是正当的，这却不足以使得这个行动是正当的。有良知的党卫军军官尽管认为其行动是正当的，或者是其义务，但通常做出正当的行动。

即使我们拒绝这种观点，这点看来还是清楚的

(F) 在大多数情形中，若某人以自认不当的方式行动，这就使得这个行动是可责备的。

在所有使得行动可责备的事实之中，这可能是最重要的。然而，在某些情形下，人们做出的事情之所以自认为是不当的，是因为他们尚未充分意识到：其行动并非不当的，而是为道德所要求的。例子之一就是哈克贝利·费恩^①帮助在逃的奴隶逃脱。这样的行动可能不是可责备的。但在大多数情形下，行动的可责备性依赖于行动在信念相对和道德信念相对的意义上是否不当。

我们接下来要追问，当我们努力决定要做什么时，“应当”、“正当的”和“不当的”的哪些意义是最重要的？在我们所讨论的这些情形以及其他许多情形中，我们行动的正当性取决于其效果或可能效果的善性（goodness）。人们通常假定

159

(G) 在这样的情形中，我们应当努力去做这样的行动，它们应当在事实相对的意义上是正当的，因为这样的行动会使得事情进展最好。

在我的医疗例子中，(G) 具有可接受的含义。我在努力做救你命的事情时，我是努力地使得事情进展最好。但是在许多其他情形中，G是错误的。考虑

矿井情形：一百个矿工被困井下，洪水正在上升。我们在地表正试图救他们。我们知道所有人都在一个矿井中，但我们不知道是在两个井中的哪一个。我们可以遥控关闭三个防洪闸。结果会是这样的：

		矿工在	
		矿井A	矿井B
我们关闭	闸门1	我们救100人	我们救不到人
	闸门2	我们救不到人	我们救100人
	闸门3	我们救90人	我们救90人

^① Huckleberry Finn, 马克·吐温的名著《哈克贝利·费恩历险记》中的主人公。

接下来我们假定：基于现有证据且正如我们所认为的，矿工在井A和井B的概率是一样的。我们若关闭闸门1或2，则有一半的机会由于能救所有的100人而做在事实相对的意义正当的事情。如果我们关闭闸门3，我们就没机会在这个意义上做正当的事情。但这显然是我们应当做的，因为我们关闭闸门3就肯定能救90个人。

160 当我主张我们应当关闭闸门3时，我是在日常意义上使用“应当”的。因为正如我们有证成地相信的，这一百个矿工有同样的概率在任何一个矿井中，所以这个行动也是我们在更精确的信念相对和证据相对的意义应当去做的。既然我们努力做在事实相对的意义正当的行动（关闭任意一个其他的闸门）是不当的，我们就应当拒绝（G）。基于一种正确观念的精炼陈述，我们可以称之为

期望主义^①：当某个行动的正当性取决于这个行动的效果或可能效果的善性时，我们应当做或努力做的事情是其结果在期望上是最好的（expectably-best）。

我们称某个行动是“在期望上最好的”，所指的并不是我们期望这个行动产生最好的结果。在这个例子中，如果我们关闭闸门3，其结果是期望上最好的，尽管这个行动肯定不会产生最好的结果，反而是关闭其他两个闸门之一可能产生最好的结果。我们为了决定哪个行动会使得事情在期望上最好，我们既要考虑不同行动之结果的善性如何，也要考虑在我们既定的信念和现有证据下，这些行动具有这些效果的概率如何。若重要的只是所救人命的数目，且某个行动是能救最大期望值的生命时，它就是期望上最好的。某个行动能救人命的期望值就是，这个行动能救的人数乘以这个行动救这些人命的概率。例如，在矿井情形中，如果我们关闭闸门1或2，所救人命的期望值就是100乘以二分之一或0.5，这个数值是50。如果我们关闭闸门3，这个期望值是90，因为这个行动肯定能救90个人。

我们同样能主张，只要我们不知道我们的行动会具有何种效果，行动效

① 这里的期望主要指的经济学上的那种期望（指某种算法下的预期），而非我们日常意义上的期望或期待。

果的期望善性就大略是：这些可能后果的善性乘以这个行动具有这些效果的概率。期望主义适用于一切情形，其中包括我们知道哪个行动事实上会使得事情最好的那些情形，这个行动的结果就是期望上最好的。

我刚拒绝了这样一种观点：当我们不知道行动具有何种效果时，我们应当努力去做事实上使得事情最好的行动。人们有时候主张，我们如果拒绝这种观点就无法解释：我们为什么在大多数情形中应当努力发现更多的事实，这样我们能够做出更有见识的决策。但这个主张是错误的。我们应当总是努力得到更多的信息，只要这种做法本身会使得事情在期望上最好。在重要的情形中，这通常是对的。在矿井例子中，如果我们能够轻易地找出矿工在哪里，努力这样做就会使得事情在期望上最好，因为我们极有可能救下所有人。

161

我们还有一个理由可以解释：当我们努力决定做什么时，我们为什么可以忽略事实相对意义上的“应当”、“正当的”和“不当的”。我们不可能努力去做事实相对意义上而不是信念相对意义上正当的事情。假定我认为，我要救你的命就必须要以某种方式行动。尽管我知道我的行动可能是错误的，我没办法努力去做事实上会救你命的行动，只能做我现在相信会救你命的行动，因为我现在相信的正是：这样做事实上会救你的命。我们无法基于事实而做决策，只能把决策基于我们现在相信是事实的那些信念上。正如西季威克同样指出的，尽管我们知道我们的道德信念可能是错误的，但我们无法去做真正正当的事情，只能去做在行动当时我们相信是正当的行动。

我之前就主张，当询问某个行动是否可责备或行为者是否有理由悔恨或愤慨时，最重要的就是这个行动在信念相对或道德信念相对的意义是否不当。我刚主张的则是，当我们在不同的行动之间选择时，不需要询问我们在事实相对的意义应当做什么。当行动的正当性取决于其效果的善性时，我们应当努力去做不是事实上会使事情最好的行动，而是去做“基于已有证据或我们既有的信念使事情在期望上最好”的行动。这些主张看起来似乎蕴含着：何种行动在事实相对的意义上当或不当根本不重要。

然而，我们可以主张：这些事实相对的意义在某种方式上是根本的。在某些实际的情形中，我们除了问某个行动是否正当之外，我们还能问更宽泛的问题：何种道德信念是真的，我们应当接受或试图遵循哪种原则或理论。我们应当努力回答其中某些问题，或至少思考其他人的某些答案。我们尽管无法去做事实上正当的事情，只能去做我们现在认为是正当的事情，还是应当努力具有真道德信念，这样做不当事物的概率就会更低。

在努力回答这类问题时，我们最好是分两阶段进行。我们首先询问，如果我们知道一切道德上相关的事实，那么何种行动会是不当的。在这样的情形下，这些问题是关于何种行动在我所称的日常意义上是不当的。但这些问题也是关于何种行动在事实相对的意义上是不当的。如果这些行动在日常意义上是不当的，且我们又知道所有相关的道德事实，那么它们也在事实相对的意义上是不当的。

回答这些问题后，我们就能转向这样的问题：当我们尚未完全知道相关事实时，我们在道德上应当做什么。这些问题是有关我们应当如何应对风险和不确定性的，因而是极为不同的。尽管正如在非道德决策的情形中一样，这些问题具有很大的实践重要性，但它们不是那么的根本。这些并不是不同的人与不同的理论具有最深刻分歧的问题。由于这两组问题具有这样的差别，我们最好分开讨论它们。因此，我通常会假定，在我的虚构情形中，每个人都知道所有相关的事实。然后我们就能追问，在这种最简单且事实相对的意义，我们应当做什么。在许多其他的情形中，这些区分并不重要，这样我通常会用“最好的”意指“最好的或在期望上最好的”。

关于这些意义与其他类似意义的“应当”和“不当的”之间具有什么样的关系，可说的还很多。例如，还有这种很难的问题：对于某人应当做什么，具有不同信念或认识到不同证据的人何时或如何会产生分歧。我的目标一直只是主张，我们需要区分这些意义，并且决定何种意义对于我们所追问的那类道德问题是最相关的。

接下来我们可以简短地回顾如下问题：我们在非道德意义上应该或应当做什么。我们现在可以说，这些问题是关于我们在实践上应当做什么。我们可把某个可能的行动称为

在事实相对的意义上是我们在实践上应该去做的事情，正好当且因为这个行动是我们具有决定性的理由或最有理由去做的事情。

“应当”的事实相对意义就是我所称的蕴含决定性理由的意义。当我们考虑“人们知道给予理由的所有相关事实”的情形时，我们使用这种意义的“应当”就够了。然而在许多情形下，人们都是不知道这些相关事实或者是具有错误的看法。在这样的情形下，我们称此行动

在证据相对的意义上是我们在实践上应当去做的，正好当我们相信

的是已有证据给予我们决定性理由要相信的，且这些信念为真，则这样的行动是我们具有决定性的理由要去做。

我们同样能够称某个行为

在信念相对的意义上是我们在实践上应当去做的，正好当我们若对这些事实的信念为真，则这个行动会是具有决定性理由去做的。

我们还能称某个行为

在规范信念相对的意义上是我们在实践上应当去做的，正好当这个行动是我们相信我们在实践上应当去做的，或者是我们相信我们具有决定性理由去做的事情。

我们除了可以在这四种意义上询问我们应当做什么，我们还可以询问何种行动是合理的。我已经主张，我们应当使用“合理的”和“不合理的”这种语词来表达赞扬与批评。理性问题在这些意义上类似于可责备性的问题。出于同样的理由，答案取决于行为者的信念。基于我早前所捍卫的这种观点：

(H) 当行动在信念相对或规范信念相对的意义上是我们在实践上应当要做的事情时，我们在理性上就应当这样做。

例如，在愤怒的蛇这个情形中，由于你错误地相信逃跑会救你的命，在理性上你就应当逃跑。在有些情形下，某个行动相对于我们关于给予理由的事实信念是合理的，但相对于我们的规范信念则是不合理的，反之亦然。

按照某些作者来说，我们在理性上是否应当做某个行动，只取决于我们的规范信念。只要我们没有回应我们认为具有决定性的理由，或者是没有做我们认为我们应当做的行动，我们就是不理性的。这类似于这样的观点：仅当行为者认为他们是不当的时候，他们才是可责备的。我认为，这样的观点太过狭隘。正如有良知的党卫军军官情形那样，即使行为者认为他们是正当的，他们的行动也可能是可责备的。我们应该类似地主张，如果我们认识到这样的事实——该事实给予我们清楚且强烈的决定性理由以某种方式行动，那么即使并未相信

这些事实给予我们这样的理由，我们在理性上还是应当这么做。这也同样适用于我们的欲望和目标。当斯加力宁愿要周二的极度痛苦也不愿要平时的一般痛苦时，这种偏好就是不合理的，即使他回应了他所相信的那些理由。

22 其他类型的不当

我刚刚说过，我们应该区分几种道德意义上的“应当”、“正当的”和“不当的”。我使用一种单一的、我所称的日常意义来定义其他几种意义。我们现在询问是否能解释这种意义，并且是否只有一种这样的意义。

我们应该主张什么是某个词的一部分意义，这可能是不清楚或不确定的。例如，“猎豹是有爪子的猎食者”是否属于“cheetah”这个词的意义，或者“必须宣战”是否属于“war”这个词的意义，这些都是不清楚的。如果决定将更多内容纳入我们对词语的意义说明，我们更多的时候就会主张一词多义。例如，我们可能会主张，war这个词具有两种意义，有一种就仅仅适用于宣告了的战争。我已经区分了几种意义的“不当的”，现在我还将区分出几种。然后我会回到一种不同的论说，它认为“不当的”只有一种道德意义。两种论说都是值得考虑的，但我们没有必要非得二者择一。

165 尽管我将探讨英词单词“wrong”，但我们的问题是关于“不当的”这个概念，即这个英语单词的意指，以及其他语言中意义足够类似的单词的意指，即这个概念指称的不当性这一属性。（当我们主张某个单词、短语或概念指称某种属性时，我们并不因此主张有任何事物具有这种属性。有许多属性是没有任何事物具有的，例如作为一种希腊神的属性，或者女巫的属性。）如果存在不同意义的“不当的”，这些意义就指称不同版本的不当的这个概念，指称不同类型的₁₆₅不当性。

就像理由的概念以及蕴含决定性理由的“应该”和“应当”概念一样，至少不当的这个概念有一个版本是不可定义的，意思是说它无法用其他术语得到有益的解释。我们可以使用这个概念来定义其他的某些道德概念。我们能说某个行动是

正当的或道德上允许的，当这个行动不是不当的时候。

以及某个行动是

我们的义务或道德上要求的，或者我们在道德上应当要做的，当我们不这样做是不当的时候。

反过来，我们也可以通过诉诸上述其他概念之一的某种不可定义版，来定义“不当的”概念的这个版本。我们也许会说，当我们不应当做某个行动的时候，这样做就是不当的。不过，虽然我们能解释这些概念是如何彼此相关的，但这个概念群都有某种共同的成分，它是我们无法单纯用语词来帮助解释的。就像理由这个概念以及应该或应当这个蕴含决定性理由的概念，这些道德概念必须让人们去思考某些想法，从而以其他方式解释。为了表达不当的这个概念的不可定义的版本，我会用“绝不可做”这个术语。

我将假定，这些道德概念也具有其他可定义的版本。例如

在可责备性的意义上，“不当的”意指“可责备的”。

在反应性态度的意义上，“不当的”意指“一类行动会让行动者有理由感到悔恨或内疚，并让其他人有理由愤慨或怨恨”。

在可证成性的意义上，“不当的”意指“无法对他人证成的”。

166

在神令的意义上，“不当的”意指“为上帝所禁止的”。

这些意义可以经组合而形成更复杂的意义。例如，当我们主张某个行动是不当的时候，我们可能意指该行动无法向他人证成，因而是可责备的；也可能意指该行动是上帝所禁止的，因而它是绝不可做的。

有些人会在蕴含理由的意义上使用“道德上应当”和“不当的”。在我们可以称之为决定性理由的意义上，

“我们在道德上应当做的事情”意指“我们有决定性理由要做的事情”。

以及

“不当的”意指“我们有决定性理由不去做的事情”。

这些意义是误导性的，不应予以采用。我们常常认为我们有决定性的理由

去做某事，尽管我们并不认为我们在道德上应当这样做。如果理性利己主义者使用这些决定性理由的意义，他们会主张

(I) 我们在道德上应当去做任何最有益于自己的行动。

但是我们最好不要把理性利己主义看作一种道德观点，而是看作道德的外部对手。按照这种观点，我们总是有决定性的理由做任何最有益于自己的事情，不管它们是否道德上不当。

在我们可称之为决定性的道德理由的意义上，

“我们在道德上应当做的”意指“我们具有决定性的道德理由要做的”。

以及

167

“不当的”意指“我们具有决定性的道德理由不去做的”。

我认为，这些意义并不是很重要。我们已经具有了“我们具有决定性理由要做的事情”这个概念，再主张其中某些理由是道德理由并无什么助益。我们应该称何种理由是“道德”理由也是不清楚的。例如，我们促进他人福祉的理由是否要全部称为道德理由就是不清楚的。这样的主张是不可能有助益的：我们道德上是否应当以某种方式行动，取决于我们应当如何答复这种在部分程度上是语词的问题。

在我们能称为道德上的决定性理由的意义上

“我们在道德上应当做的”意指“我们具有道德上的决定性理由要做的”。

以及

“不当的”意指“我们具有道德上的决定性理由不去做的”。

尽管“道德上的决定性理由”看来非常类似于“决定性的道德理由”，但它们之间有两个重大的差别。第一，当询问我们是否具有道德上的决定性理由以某种方式行动时，我们询问的不是我们是否具有那种应该称为“道德的”理由，而是我们要这样做的理由是否在道德上压倒所有不这样做的理由。第二，为了能够在决定性的道德理由这个意义上使用“不当的”来陈述我们的道德信念，我们必须相信我们总是有决定性的理由不做不当的行动。但是如果我们反过来主张，我们具有道德上的决定性理由不如何行动，那么这依然无法决定这些理由是否在非道德意义上也是决定性的，或者是综合来看是决定性的。即便我们相信在有些情形下，我们有充分的或决定性的理由去做不当的行动，我们依然能在这种意义上使用“不当的”。

有些人看来使用

“我们在道德上应当做什么”意指“我们具有最强的不偏不倚的理由要做什么”。

某个行动在上述意义上是不当的，如果我们具有更强的不偏不倚的理由去做其他事。我们能称这些是在蕴含不偏不倚的理由意义上的“应当”和“不当的”。正如我所说，还有着类似意义的“好的”、“坏的”和“最好的”。按照某些行动后果论者的观点：

我们的行动应当总是使得事情变得最好。

如果这个主张是在蕴含不偏不倚的意义上使用“应当”和“最好的”，它的意思就会是：

(J) 我们具有最强的不偏不倚的理由要做的，就是使事情按照如此方式进行的任何行动——该方式是我们所有人都有最强的不偏不倚的理由想要事情那么进行的。

我们可称此为不偏不倚的行动后果论。为表达这种意义上的应当，我们可以使用“不偏不倚的应当”这个术语。

这种意义上的“应当”与更为常见的道德意义颇为不同。例如，西季威克

写道：

“从宇宙的视角看，任何单个人的利益并不比任何其他人的利益更为重要……作为一个理性存在者，我的目标必定在于普遍的利益……而不仅仅在于其中的某个特定部分……我不应当偏爱自己较少的利益而不是他人较多的利益。”

当西季威克主张他不应当偏爱自己较少的利益时，他的意思似乎不是这种偏好是可责备的，或者无法向他人证成的，或者这个行动会让他有理由悔恨和让他人有理由愤慨。西季威克的意思似乎只是，从不偏不倚的视角来评估，他给予自己较少利益的理由要弱于他给予其他人更多利益的理由，或者说前一种理由被后一种所压倒。

169

这类后果论最好被看做是如理性利己主义一般的、道德的外部竞争对手，而不是一种道德观点。鉴于这种观点主张我们应当牺牲己之小利来成全他人之大利，它就离道德要近得多。这使得该观点在某些方面成为更值得认真对待的对手。许多拒绝理性利己主义的人可能会接受不偏不倚的行动后果论，因为他们认为自己的福祉是西季威克所称的“狭隘的”或“卑鄙的”目标。

(J) 可能看来是一个无足轻重的主张，近乎同义反复。然而，如下主张并非无足轻重的：行动在这些蕴含不偏不倚理由的意义上可有正当与不当之分，结果可有好坏之分。如我曾经主张的，基于某些广为接受的理由观，并不存在任何这样的行动或后果。而且即使(J)是同义反复的，不偏不倚理由的行动后果论也能够提出其他的实质性主张。例如，这些人如果是享乐主义的效用主义者就可以主张

(K) 我们应当不偏不倚地去做的事情要能产生最大总量的幸福。

这些人可能认为，我们都有强有力的理由这样做。他们的行动可能不会基于那些涉及更为常见意义的“道德上应当”和“不当的”的道德信念，甚至根本就不具有这样的信念。这些人可能深信，“事情进展如何”关系重大，他们可能有强烈的动机和常被打动去阻止或减轻痛苦。但他们可能困惑于某些行动是否是义务或者是绝不可做，也困惑于可责备性和有关悔恨和愤慨的理由。这在一方面表明，这种形式的后果论是道德的外部对手。

如我所言，按照某些作者的观点，“应当”、“正当的”和“不当的”仅有一种道德意义。对于诸多可定义的意义来说，做出这种主张显然是讲不通的。如果我们可以在一种可定义的意义使用“不当的”，我们无疑可以在其他可定义的意义使用它。我们同样无法言之成理地主张：存在每个人都使用的唯一且可定义的一种意义。我们甚至都无法主张：每个人使用“不当的”是意指“我们具有道德上的决定性理由要不做的”，因为有些人从不使用或很少使用理由这个概念。

更言之成理的是这样的主张：每个人都是在我以“绝不可做”这一短语所表达的这种不可定义的意义使用“不当的”。我们也许还可以主张，可责备性和反应性态度的意义暗暗地诉诸这种不可定义的意义，因为责备、内疚、悔恨和愤慨的态度都涉及这样一个信念：某个行动是不当的。在用“道德上的决定性理由”定义“不当的”的意义时，我们可能不得不对“道德上”这个词作不可定义的使用。人们可能会主张某个其他的可定义的意义要表达的不是“某个行动是不当的”这个信念，而是有关不当行动的其他信念。例如，神令意义和可证成意义可能是在如下意义上来表达“行动是不当的”这个信念：当且因为这些行动是为上帝所禁止的或者无法向人们证成的，那么我们绝不可做它们。

当某个作者主张像“不当的”和“应当的”这样的词只具有一种道德意义时，他们诉诸的事实是：即使当我们和其他人具有极为不同的道德观点时，我们也只是把自己看做是与这些他人的观点发生分歧。这些作者主张，如果我们和其他人是在不同的意义上使用这些语词，我们就不可能与这些人具有分歧，因为我们不是在讨论同一个问题。

这个论证是软弱无力的。不同的人对“不当的”和“应当”可能是在可定义的意义使用的，这些意义是不同的但有部分是重叠的。这就足以使得分歧成为可能。例如，假定当我主张某个行动不当时，我的意思是，此类行动是上帝所禁止的因而是可责备的。当你主张某个行动不当时，你的意思是，这类行动无法向他人证成因而是可责备的。如果我主张某个行动是不当的，而你主张它不是不当的，我们可能会在这个行动是否可责备上具有分歧。当人们在这类不同的意义上使用“不当的”时候，这可能会增多他们的分歧。在刚才的虚构情形中，如果我们理解彼此对“不当的”用法，你就可能会认为，没有任何行动在我的意义上是不当的，因为你认为没有任何行动会因为它们是上帝所禁止的而是可责备的。而我则相信没有任何行动在你的意义上是不当的，因为我不认为任何行动会因为它们无法向他人证成而是可责备的。我们的分歧由此是

彻底的，因为我们每个人都完全拒绝另一个人的所有道德信念。

171 当同一共同体中的人在不同且部分重叠的意义上使用“不当的”和“应当”这种语词时，这些人有理由转向其他更薄的意义，这是他们都可以使用的。由此这些人何时何处会产生分歧，就会是更显而易见的。在刚才的虚构情形中，如果你和我都用“不当的”意指可责备性，我们就能同意许多行动在这个意义上是不当的，即使我们在什么因素使得行动不当上有所分歧。

我们可以补充一点：在某些情形中，那些在不同的意义上使用“不当的”和“应当”的人有时候可能没有分歧。例如，基于西季威克的观点，如果牺牲自己的生命能够救两个在相关方面与我差不多的陌生人的命，我就应当这样做。如果西季威克是在可责备性或反应性态度的意义上使用“应当”，我们大多数人都会拒绝这个主张。我们会认为，如果我救自己的命而不是两个陌生人的命，我的行动无可指责，我没有任何理由为之悔恨，不管是这两个陌生人还是其他人都没理由为此而愤慨。但西季威克的意思可能只是，我具有更强的不偏不倚的理由来救这两个陌生人，而这个主张不会与其他人的道德信念相冲突。

我们接下来考虑这样的情形，其中行动的正当性取决于其效果的善性。在这样的情形中，有些人主张

(L) 我们的行动应当使得事情最好。

其他人则主张

(M) 我们的行动应当使得事情在期望上最好。

如果(L)在事实相对的意义上使用“应当”，而(M)在证据相对的意义上使用“应当”，这两个主张就并不冲突，且我们都能接受。这两个主张都不会冲突于(M)某一版本，该版本在信念相对的意义上使用“应当”和“在期望上最好的”。

我们还有一种分歧是可避免的。按照某些作者的观点，我们应当做某些事情，诸如信守诺言、救人性命和使得事情在期望上最好。按照其他某些作者的观点，我们应当尽力去做这些事情。我认为，我们应当同时做出这两种主张。我们不应该只是主张我们应当做这些事情，因为我们是否尽力去做这些事情在道德上是重要的。例如，如果我们尽力了但没有做到遵守诺言，我们并不应得

任何责备。我们也不应该只是主张我们应当尽力去做某些事情，因为我们成功与否在道德上往往是重要的。例如，如果我们信守诺言的企图失败了，那么情况可能是我应当转而以其他方式行动。我们若主张应当做某件事情，就常常被看作意指我们应当做这件事情，或至少要尽力去做。

我刚才所描述的各种意义，它们是否应当被称作是“不当的”的不同意义（指向不同类型的正当性），这是不重要的。我们区分这些意义以及它们所表达的概念，这就足矣。我们由此能决定何种概念是最值得使用的。

我们在做出这种决策时可回到这个问题：道德在蕴含理由的意义上有多重要。如果某个行动是不当的，这个事实就给予我们理由不做它吗？如果是这样，那么理由有多强呢？

这个答案在部分程度上依赖于我们用“不当的”意指什么，依赖于我们使用“不当的”指称何种正当性。首先假定，我们若主张某个行动不当，意思就是我们具有决定性的道德理由不这样做。这些理由是由“使得这个行动不当”的那些事实所提供的。两个这样的例子是：某个行动是撒谎的事实或会引起无意义的痛苦这个事实。基于这种观点，这个事实

(N) 某个行动是不当的。

会是下述事实的高阶事实

(O) 存在着其他的事实会给予我们决定性的道德理由不这样做。

这种高阶的事实不会给予我们进一步且独立的理由不这样做。尽管我们可能主张一个行动的不当性总是给予我们理由不去做它，但这个理由可能是派生的，因为它的规范效力可能完全派生自给予理由的其他事实。这样如果仅在这种“决定性的道德理由”之意义上使用“不当的”，我们会主张：

(P) “某个行动若是不当的”这个事实不会给予我们任何进一步的理由不这样做。

按照这种观点，某个行动是否不当就不具有任何实践重要性。当我们努力决定要做什么时，我们只要询问“我们是否有决定性的理由赞成还是反对做某

个可能的行动”就足矣。我们若判定自己具有这种理由，就可以询问这些是否是道德理由，由此我们的行动在决定性的道德理由之意义上是否不当。但这不是我们应当做什么或有理由做什么的问题。但这个问题仅仅只是概念性的，就像何种理由最好被称为法律理由或美学理由那样。因此，我认为，我们没有多少理由要在这种意义上使用“不当的”。

许多人假定，行动的不当性确实给予我们强大甚至是决定性的进一步理由不去做它。如果这些人是在“决定性的道德理由”之意义上使用“不当的”，这些假定将是错误的，其错误方式正如我刚才所描述的。这并不表明：这些人可能由于没有明白我刚才所提出的观点，因而不能在这个意义上使用“不当的”。但我认为，大部分人是在一种或更多种其他的意义上使用“不当的”的。当某个行动在其他的意义上是不当的时候，我们就能主张，这些行动的不当性给予我们进一步且独立的理由不那样去做。

首先假定我们在不可定义的意义使用“不当的”。我们若主张某个行动在这个意义上不当，则并非主张这个行动具有斯坎伦所称的“具有给予理由的其他属性”的那种“纯粹形式且更高阶的属性”，而是主张，这个行动具有“作为绝不可做的某事”这种极为独特的实质属性。尽管我强烈地相信某个行动在几种其他的意义上是不当的，但是，是否有任何行动具有这种不可定义的属性，在我看来这是个悬而未决的问题。如果确实如此，我们就能言之成理地主张，某个行动若绝不可做，这给予我们强有力的理由不这样做。这是“不当的”意义之一，我们能最言之成理地用它来主张：

(Q) 某个行动若不当，这个事实总是给予我们决定性的理由不去做它。

即使我们用“不当的”意指“为上帝所禁止的”，这尽管是出于极为不同的理由，也丝毫不会削弱(Q)的可行性。

174 如果我们在其他可定义的意义使用“不当的”，我们同样能主张，行动的不当性给予我们独立的理由不去做它。某个行动若可责备，无法向他人证成，给予理由让我们悔恨和让他人愤慨，这些事实全都能给予我们进一步的理由不去做它。然而，我们不应该主张，这些事实将总是会给予我们最强的理由不去做不当的行动。例如，如果某个行动会引起巨大的痛苦，这个事实给予我们的理由，会比这个行动可责备或无法向他人证成这种事实所给予的理由要强大得多。

正如我讲过的，我们没必要在“不当的”这些意义以及它们所表述的概念之中做出选择。例如，在如下提问之中使用其中若干概念将是有价值的：何种行动是在不可定义的、不可证成的、反应性态度的或可责备性的意义上是不当的。在本书的余下部分，我会在组合的意义上含糊地使用“道德上应当”和“不当的”。

我们应该如何理解这些规范性的概念，行动是否能具有这些概念所指称的属性，对此有着更深刻且困难的问题。除开第六部分，我将极少谈论到这类元伦理的问题。一旦我们在实践理由、认知理由和道德的思考上取得更多的进展，这些问题就更易于回答。正如罗尔斯和内格尔所主张的，我们的道德理论“是原始的，极为不足”，“伦理理论……还处于婴儿期”。

我不是要提出一种新的道德理论，只是努力从某些现存的理论之中学习，以望离真理更近一点。我们将从康德开始，因为他是自古希腊以来最伟大的道德哲学家。康德在提出其著名公式时写道，他的目标是找出“道德的最高原则”。我将追问他是否成功了。

第二部分

各种原则

第八章 可能的同意

23 强制与欺骗

按照康德最受人喜爱的原则，通常被称为：

177

人性公式：我们必须把所有理性存在者或人（persons）永远不要纯粹作为手段，而要总是作为目的来对待。

康德主张，要把人作为目的，我们就永远不能以其不能同意的方式对待他们。例如，康德在解释许假诺的不当时写道：

若我想用假诺言来利用某个人，从而实现自己目的，则这个人不可能同意我用这种方式对待他。

科丝嘉评论道：

人们若得不到任何机会对某做法表示同意就不可能同意该做法。最明显的这类场合就是当人们使用强制的时候。但这也同样适用于欺骗……知道会发生什么且对进程有所掌控，是有可能同意的前提条件。

奥尼尔同样写道：

如果我们强制或欺骗他人，那么他们的不同意，由此他们的真正

同意也就从原则上被排除在外了。

178

科丝嘉得出：

按照人性公式，强制和欺骗是不当地对待他人的最根本的形式。

这些评论显示了这种论证：

以人们不可能同意的方式对待他们是不当的。

人们不可能同意受到强制或欺骗。

因此

强制和欺骗就永远是不当的。

然而，有时候以人们不可能同意的方式对待他们是正当的。例如，当人们昏迷时，他们不可能同意救其生命的手术，但这并不会使得这种手术是不当的。

科丝嘉可能会说，康德的主张只适用于“其性质使得同意不可能”的那种行动。不同于手术，欺骗就是这样的行动。人们要能同意我们对待他们的方式，就必须知道我们在做什么。如果人们知道我们正在欺骗他们，那么我们就不能欺骗他们。这样我们就不能在人们同意的情况下欺骗他们。这就是为什么不像手术那样，欺骗总是不当的。但是考虑

致命的信念：我知道，除非我对你撒谎，否则你就会确切地知道布朗是个谋杀犯。由于你无法在布朗面前掩饰这个信念，他也会谋杀你。

179

如果我什么也不说，那么你在临死的时候就可以合情理地抱怨，我应该欺骗你来救你的命。我将无法有证成地这样答复：由于我不能经你同意而欺骗你，所以救你命的这种方式是不当的。我的救命谎言就会如同对昏迷者的救命手术。正如这个人若能就会同意这种手术一样，你也会同意我欺骗你。如果我征询你的同意，这会使得我不可能欺骗你，但这只是一个技术问题。如果你有能力使得你失去特定记忆，那么我们就解决这个技术问题。在你给出你的同意之后，你能有意地遗忘这段对话，由此我的撒谎就能救你的命。既然如果你

能就会同意我欺骗你，那么我的撒谎在道德上就是无罪的，就像为了给某人一个意外派对而有必要撒谎一样。

这也同样适用于强制。我们也许会主张，人们无法同意受到强制，因为一旦人们给予了同意，那么他们就不可能是受强制，如果他们是受到强制的，他们就无法给出同意。但是我们能自由地同意以后受到某种方式的强制。在发现麻醉剂之前，许多人都同意在随后痛苦的手术过程中受到强制。即使我们正受强制，我们也能自由地同意某类强制。我们大部分人都会投票，同意以惩罚之威胁来支持公平的税收和对良法的服从，从而让每个人一直受到法律强制。为了改善混乱无序的状况，我可能同意受强制。尽管强制和欺骗通常是不当的，但我认为，使得它们不当的不是这个事实：这些行动的性质使得同意是不可能的。

24 同意原则

我们现在回到康德的这个主张：

(A) 以人们无法表示同意与否 (can't possibly consent) 的任何方式对待他们都是不当的。

科丝嘉写道，“当人们根本没有机会同意时”，他们就不可能同意。奥尼尔同样写道，“要把他人当做人来对待就必须给予他们这样的可能性：能对所提议的事情表示同意或不同意”。这些评论假定，康德的意思是

(B) 以人们无法表示同意与否的任何方式对待他们都是不当的，因为我们没有给予他们给予或是拒绝同意的可能性。

当我们以某种方式对待人们时，他们通常告诉我们或他人他们是否同意，由此在宣告的意义上给予或拒绝同意。科丝嘉和奥尼尔是在一种不同且更重要的意义上使用“同意”。人们如果具有科丝嘉所称的“对事情进程的掌控力”，那么就能在这种影响行动的意义上给予或拒绝同意，因为只有他们同意才能受到某种方式的对待。这样我们能够把 (B) 重述为

180

给予选择的原则：如果我们不给予他人以得到何种对待的选择能

力，这是不当的。

如果这就是康德的意思，那么我们就不得不拒绝康德的主张，因为给予选择的原则显然是错误的。例如，这个原则错误地蕴含了我们应当给予其他人选择：我们是否给他们的学生打低分、买他们推销给我们的东西、要回他们从我们这里偷走的东西、报告他们的罪行或在选举中投票反对他们。不仅如此，在道德上重要的大部分情形中，我们对不同可能行动的选择会对不止一个人产生重要的影响。只要两个人会做出相冲突的选择就能表明，我们就不可能给予这两个人选择我们如何行动的能力。这样一来，给予选择的原则也错误地蕴含：在上述所有情形中，不管我们做什么都是不当的。

我认为，对康德的观点我们有着更好的阐释方式。科丝嘉和奥尼尔都假定，当康德主张

(A) 以人们无法表示同意与否的任何方式对待他们都是不当的。

他意指

(C) 由于我们没有给予人们能力选择受到何种对待，故他们不可能在影响行动的意义同意我们对待他们的方式，因此这种方式就是不当的。

我认为，康德的意思是

181

(D) 如果我们若给予人们能力选择受到何种对待，则他们在影响行动的意义本不会同意我们对待他们的方式，那么这种方式就是不当的

人们也许会有这样的反驳：如果我们给予人们这种能力，他们就能够选择我们的行动方式，那么就不存在着任何行动是这些人不可能同意的。如果这就是康德心中所想的那种不可能性，(D) 就是无足轻重的，因为 (D) 就从未蕴含着某个行动是不当的。但是还存在着另一种不可能性。人们若说“我无法对你的建议表示同意与否”，极少意指同意不是他们能做出的选择之一。这些人

几乎总是意指他们无法理性地同意，因为他们有决定性的理由拒绝同意。我认为，康德的意思是

(E) 以人们无法理性地同意的任何方式对待他们都是不当的。

我称此为同意原则。

有几个理由认为康德诉诸的正是这个原则。虽然给予选择的原则显然是错的，但同意原则可能是对的，这使得它更有可能是康德的本意。当康德主张我们不能做某事时，他常常意指我们不能理性地做这件事。康德还写道，如果他不当对待某人，这个人

没有可能同意对我对待他的方式，从而他本人也不可能容纳 (contain) 在这个行动的目的之中。

如果康德是在主张我们应当让别人选择我们对待他们的方式，他就没有理由加上这一点，因为我们对待某人的方式能得到证成，这个人必须能够通过分享这个行动的目标，从而“容纳在这个行动的目标之中”。一旦我们让他人选择我们对待他们的方式，我们的行动就不带有这些人不可能分享的目标。康德的意思必定是，当我们选择如何对待他人时，我们的行动应当总是具有这些人能够分享的目标。这些人能分享我们的目标，这一点即使是可设想的也是不够的，因为许多无法证成的目标也能被设想成是可分享的。我们的行动应当只具有其他人能够理性地分享的目标，这样他们就能理性地同意我们对待他们的方式。

182

尽管康德有关目的或目标的言论有助于解释他有关同意的主张，但这些言论自身却不那么讲得通。即使他人能够理性地分享我们的目标，但若这些人不能理性地同意我们实现这些目标的方式，我们的行动还是可能不当。尽管你可能理性地分享我驯服的老虎应该得到喂养这个目标，但不可能理性地同意自己成为老虎的食物。即使其他人不能理性地同意我们的目标，但若他们能理性地同意我们的行动，则我可能就没有行动不当。尽管你不可能理性地分享引用某个人的名字一千次，你却可能理性地同意我引用你的名字。这样一来，相比于其他人是否能理性地分享我的目标而言，人们是否能理性地同意我们的行动是更为重要的。

康德有关同意的主张给予我们一个鼓舞人心的理想，即我们作为理性存

在者应当彼此如何关联的理想。我们是否能够实现这个理想是值得追问的。我们不可能总是让每个人选择我们要如何对待他们。但是 we 也许能以人们能够理性地同意的方式对待他们。如果这是可能的，康德的主张“此即每个人总是应当如何行动的方式”就可能是正确的。

25 给予同意的理由

我们能否实现康德的理想，这要取决于人们能够理性地同意的是何种行动。罗尔斯表示，康德在提出同意原则时假定

(F) 人们能够理性地同意某个行动，当且仅当或正好当，他们能够意愿行为者的准则 (maxim) 的确成为普遍法则。

罗尔斯在此参考了康德对道德的最高原则所提出的另一种陈述。按照康德的

183

普遍法则公式：按照我们无法意愿其成为普遍法则的任何准则行动都是不当的。

对于我们的准则，康德大略意指我们的策略和根本目标。我们还没必要考虑准则在什么意义上可以被意愿成为普遍法则。

然而，康德本人并未承诺 (F)，所以这个假定可能是错误的。假定我是你的医生，我问你是否同意我给予你某种治疗。你的同意要想是合理的，你需要相信我是一个合格且有良知的医生，知道各种可能治疗方案的可能效果。但是你没必要知道，我是否基于你意愿其变为普遍法则的某个准则或策略来行事。

为了支持其提议，即康德假定了 (F)，罗尔斯诉诸康德的如下评论：他的各种原则都不过是“完全相同法则”的不同陈述而已。罗尔斯认为，这个评论蕴含着康德的其他原则“不能为康德的人性公式增添任何内容”。罗尔斯因此提出，我们应该这样来阐释康德的其他原则，使得它们无法增加任何东西，因为它们并不包含任何其他的想法。

这个提议贬低了康德作为哲学家的成就。康德本人甚至更进一步地贬低他的成就，因为他否认他甚至是在提出一个新原则。真相在于，短短四十页的《奠基》却大放光芒，其中给出的富有成效的新观念，比几个世纪以来所有的

哲学家提出的还要多。不一致性正是康德能够具有如此成就的特质之一。我们如果由于康德的某些主张与其他主张相冲突而予以忽视，那就可能错失康德观念所具有的、如赫尔曼所称的某些“未经开发的理论能量与活力”。

康德的同意原则就是此类例子之一。令人奇怪的是，这个原则甚少得到人们的讨论。这个原则具有极大的吸引力，值得作为一个独立的道德观念加以考虑，而不仅仅是作为对康德普遍法则公式的另一种陈述。因此，我在询问这个原则蕴含什么时并不假定（F）。

当我们询问某人是否能理性地同意某个行动时，我们的问题应该是有关在影响行动的意义上的那种同意。如果人们拒绝同意只不过是种单纯的宣告或抗议，对于其他人会如何对待他们毫无影响，甚至可能使得对待方式变得更糟，那么我们询问人们是否能够理性地同意受到某种方式的对待就是没价值的。如果情况真是如此，那么这些人不去抗议可能是理性的，即使他们受到对待的方式很坏或者极为不当。

我们的问题还应该是有关知情同意的。若人们不知道某个行动具有的可能效果，那么他们是否理性地同意此行动就是不相关的。如果人们不知道自己会受到何种对待，那么他们可能会理性地同意极严重的虐待。出于这些理由，我们可以把同意原则重述为

CP: 以任何这样的方式待人是当的：如果人们知道相关的事实，且我们给予他们能力来选择我们如何对待他们，那么这种方式是他们不可能在影响行动的意义理性地同意的。

我们若知道自己的行动或者其他的替代行动会或可能以某种方式影响人们，或者是他们具有个人理由要关注的，我们就可以被算作是以某种方式对待他们。即使我们的行动方式不会在因果上影响这些人，情况依然是如此。没有成功地救某人的命以及不遵守对死者的承诺就是两个这样的例子。

若人们知道相关的事实，且正好当这些事实会给予他们充分的理由去同意某个行动，则他们就能够理性地同意它。若这些理由并不弱于人们能拒绝同意的任何理由，则他们就有充分的理由去同意。这样同意原则可能更简洁地陈述为

CP2: 若人们在影响行动的意义上没有充分的理由同意对待他们的方式，则这种方式就是不当的。

我这样陈述这个原则时假定：我们只要对于理由或显见理由做出回应就是理性的。基于合理性的其他观点，CP和CP2陈述了不同的原则，它们可能会具有不同的含义。你若接受这样一种观点就应该认为，同意原则是由CP2所陈述的。当我询问某人是否能理性地同意某个行动时，我询问的是，这个人是否具有充分的理由去同意。

同意原则要想成功，就必须既是本身说得通，具有的含义也说得通。这个原则必须不能要求太多在我们看来显然不当的行动，或谴责（在蕴含的意义上是不当的）太多在我们看来显然是道德所要求的行动。这个原则若既蕴含许多我们最深思熟虑的直观道德信念，还能言之成理地支持这些信念，则我们就能有证成地使用这个原则来指导某些信念，从而修正或扩展它们。

同意原则蕴含什么，要取决于这个假设：何种事实给予我们理由。如果我们假定的是某种基于欲望的主观理论，或者是理性利己主义，那么同意原则是说不通的，它会错误地谴责许多可允许的或者道德所要求的行动。例如，假定在

地震情形中，怀特和格雷两人被困于正逐渐坍塌的废墟中。我是救援者，能阻止废墟杀死怀特或毁坏格雷的腿。

我们应该假定，我、怀特与格雷彼此都不认识，我们在道德相关方面均是一样的。我们在后面的虚构例子中也有类似的假设。如果这些就是唯一仅有的相关事实，那么很清楚我应该救怀特的命。我们接下来假定，如果我救了格雷的腿，这对格雷会好很多，会更好满足格雷当下完全知情的欲望。按照基于欲望的主观理论和理性利己主义两者来说，格雷都不能理性地同意我不救他的腿，同意原则由此就会错误地谴责我救怀特的命是不当的。这也同样适用于无数的其他情形。按照主观理论和理性利己主义两者来说，有着无数正当行动是人们无法理性地同意的。如果我们接受这样的原则，正如许多人所做的，那么我们必须拒绝同意原则。这也许是因为这个原则如此不受人关注的原因之一。

186 我已经主张，我们应当接受某种广义的基于价值的客观理论。基于这种观点，如果一个选择会使事情在不偏不倚意义上变得更好，而另一个选择会使事情对自己或与我们具有紧密纽带的人更好，我们通常具有充分的理由二者择一。我认为，地震就是这样的一个情形。如果格雷能够选择我如何行动，我认为，她就具有充分的理由做任一选择。格雷能够理性地选择我救她的腿，因为

这个选择对她要好得多。但是理性并不要求她这样做。格蕾可以反过来理性地选择我救怀特的命。格蕾可以理性地认为怀特的福祉与她自己的一样重要，怀特失去生命的损失要远大于她失去腿的损失。

相对比而言，怀特不能理性地选择我救格蕾的腿。我认为，我们通常能理性地选择有益于陌生人，即使我们的这个选择会使得自己的损失要稍大一点。但怀特与格蕾的可能获益在此有着太大的差距。怀特不会具有充分的理由放弃自己的生命让我来救格蕾的腿。因此，同意原则正当地要求我救怀特的命，因为这是格蕾和怀特都能同意的唯一行动。

接下来我们假定，在

救生艇情形中，我搁浅于一块岩石上，另五个人搁浅于另一块岩石上。在上升的潮水淹死所有人之前，你能用一个救生艇救我或救另外五个人。我们都很年轻，死亡的损失都差不多。

尽管有些人认为，你应当给我某种得救的机会（这可能是六分之一甚至二分之一），大部分人能有理性地认为，你应当救那五个人。

如果我能选择你如何行动，我能理性地选择你救五个人而不是我吗？有些人认为不能。这些人可能同意，如果我选择放弃我的生命以救五个人的命，那么这种选择是道德上可敬的。但他们认为，这个选择还是不合理的。基于这种观点，由于我无法理性地同意你救五个人而不是我，所以同意原则蕴含着你救五个人是不当的。这个结论是令人无法接受的。因此，如果接受这种观点，我们就不得不拒绝同意原则。

187

我认为，我们应当拒绝这种观点。尽管我能理性地选择你救我，但我认为，我还能理性地选择你救那五个人。如果放弃自己的命能救五个陌生人的命，那么我就有充分的理由这样做。

这五个人能够理性地同意你救我而不是他们吗？“同意”这个词可能在此是误导性的，因为我们可能假定，五个人中的每一个都只能代表自己来给予同意。但我们不应该做出这种假定。当我们应用同意原则时，我们询问的应该是，如果五个人中的每一个都能够给出在影响行动的意义上的同意或不同意，由此选择你如何行动，那么这个人是否能理性地同意你救我而不是这五个人。答案显然是否定的。我们假定格林是五个人中的一个。格林没有充分的理由选择你救我而不是既救格林也救其他四个人。格林既有强烈的个人理由也有不偏

不倚的理由不这样选择。基于这些假设，同意原则正确地蕴含着你应当救那五个人，因为这是我和五个人中的每一个都有充分的理由能同意的唯一行动。

正如这些例子所显示的，我们能否理性地同意某个行动，在部分程度上取决于这个行动和其他行动产生的不同结果给我们和其他人所带来的利益或负担。这些利益或负担有多大，它会带给多少人，都是会产生影响的。我认为，我们和其他人过得如何也会产生影响。我们和他人对于情境的各种特征是否负有责任也可能会产生影响。例如，如果我们有些人通过工作产生了这些利益，或者由于疏忽或不小心而产生了这些负担，情况就可能是如此。可能还有些行动是我们没有充分理由同意的，即使这些行动不会给我们带来很重的负担。例如，我们可能具有很强的理由拒绝同意其他人来决定我们如何生活，即使这些人的决策对我们并不是坏的。

只要人们对于受到某种方式的对待不能理性地给出知情的同意，就必定存在着有关这些行动的事实，它们会给出决定性的理由让他们拒绝同意。我已经讲过，由于怀特的损失远大于格蕾的损失这个事实，所以怀特不能理性地同意我救格蕾的腿而不是他的命。我们也能可行地主张，这个事实使得这个行动是不当的。这也同样适用于其他的情形。只要某些事实给予某些人决定性的理由拒绝受到某种方式的对待，这些事实也会对这样的行动提出道德反驳。

同意原则要为真，这些道德反驳必须是决定性的，因为这个原则谴责任何人都不能理性地同意的一切行动。这个强硬得多的主张要想能得到辩护，情况就必须总是或近乎总是

(G) 至少有一种可能的行动是每个人都有充分的理由同意的。

我们能称(G)为全体一致性条件。在不存在这种行动的情形中，同意原则会错误地蕴含着，无论我们做什么都是不当的。(G)最不可能为真，若

(H) 可能行动中的每一种会给至少一个人施加极大的负担，或剥夺至少一人的极大利益。

这些人会具有极强的理由拒绝同意承担这种负担或被剥夺这种利益。救生艇情形就是这样的一种情形：其中要么我要么其他五个人早夭。我已经讲过，在这个情形中，(G)为真。尽管我有非常强的理由选择你救我的命，但这些理

由不会是决定性的。我认为，我会有充分的理由同意你救五个人的命而不是自己的命。如果我拥有这样的理由，那就会强烈地支持这样的观点：至少在筹码不是很大的情况下，存在至少一种可能的行动是每个人都能理性地同意的。

我会回到这个问题：是否总是存在这样一种行动。如果确实如此，我们可以论证：

只要某人不能理性地同意某个行动，那么必定存在某些事实——它们给予这个人决定性的理由拒绝同意。这些事实对这个行动提出了道德反驳。

这些反驳必定比每个人都能理性地同意的其他可能行动所受到的反驳要强得多。

只要两个行动之一受到的道德反驳要强得多，那么这个行动就是不当的。

因此

一种行动方式，只要人们不能理性地同意，就是不当的。

尽管这个论证是粗略的，但它足以表明同意原则自身是说得通的。

这个原则也有许多讲得通的含义，因为它谴责了许多明显不当的行动，诸如杀人、伤害、强制、欺骗、偷盗和毁诺。许多这样对待人们的方式都是人们没有充分的理由同意的。

26 一个多余的原则？

按照某些作者的看法，诉诸理性同意的可能性将一无所成。这些人承认：以人们无法理性地同意的方式对待他们永远是不当的。但这些作者写道，更为重要的不是“那些人不能理性地同意这些行动”这个事实，而是“给予这些人决定性理由拒绝同意”的各种事实。

在考虑这个反驳时，我们可以先区分任何道德原则可能实现的两种目标。这种原则可能提供关于不当性的某种可靠标准，确切地告知我们某类行动全都是不当的。这个原则也可能是解释性的，描述这些行动为什么不当的理由之一，或

者描述使其不当的事实之一。按照我刚才提到的这些作者的看法，即使同意原则为真，我们既不需要这个原则作为一种标准，这个原则也不是解释性的。

若我们考虑的行动主要影响到一个人，我们无法与这个人交流且不知道其偏好，则这个反驳最说得通。在这样的情形中，我们还是不得不替这个人做出某个决策。例如，医生有时候不得不替昏迷的病人做出决策。在这样的情形中，这样主张就足够了：我们应当尽力去做对这个人最好的事情，或最有益于这个人的事情。若补上“我们以这个人不可能理性同意的任何方式行动都是不当的”，就纯属画蛇添足。

然而，在最重要的情形中，我们在不同行动之间的选择可能严重地影响到两个或更多的人。我刚才所描述的观点可以拓宽，从而涵盖这样的情形。按照行动效用主义，或者

AU：我们应当总是做在总体上最有益于人们的任何事情，给予人们（减去负担之后）最大的利益。

行动效用主义者会主张

(I) 每个人都能够理性地同意做总体上最有益于人们的一切事情，且仅会同意做这样的事情。

如果(I)为真，AU和同意原则就总是吻合的，它们要求的是完全相同的行动。这些效用主义者就会主张AU是更根本的。既然AU告诉了我们总是应当如何行动，同意原则对于我们的道德思维并没增加任何东西。但这个主张是错误的。如果每个人能够理性地同意的行动只有那些效用主义行动，同意原则就会支持AU。(I)为真会给予我们进一步的理由认为这些行动是道德上要求的，有进一步的理由去这样做。

191 我认为(I)并不为真。有许多效用主义行动是某些人不能理性地同意的，而有许多非效用主义行动是每个人都能理性地同意的。我将在后面给出这样的一些例子。

如果同意原则为真，它将不止是不当性的一个可靠标准。只要某人不能理性地同意受到某种方式的对待，这个事实就会给这个行动提出一种反驳，人们可以主张它是使得这个行动不当的事实之一。当我们必须在多种可能行动——

这些行动会对彼此利益或目标相冲突的许多人产生重大影响——之间做选择时，同意原则最具重要性。在这样的情形中，如果只存在一种行动是每个人都能理性地同意的，这个事实就会给予我们很强的理由这样去做，并且它本身就足以解释为什么其他所有可能的行动都是不当的。

我们还有另一个理由询问同意原则是否为真。即使我们并不需要同意原则作为不当性的标准，仍然值得询问我们是否能实现康德的理想：只以人们能够理性地同意的方式对待他们。

27 实际同意

人们是否实际上同意受到某种方式的对待，或者如果给予机会的话，他们是否会实际上同意，这在道德上通常是重要的。在这样的情形中，只问人们是否能理性地同意某个行动是不够的。强奸犯可能主张，受害者本来能理性地同意与他发生性关系。即使这个主张为真，这也不能证成这个人的行动。人们由此可能会反驳，同意原则由于并不要求实际同意，就错误地忽略了实际同意的道德重要性。

然而，这并不为真。即使这个受害者本来可能会理性地同意与强奸犯发生性关系，她也不可能理性地同意受强奸，即她实际上拒绝而被迫的性交。在这个和其他许多类型的情形中，没有我们的实际同意，我们不可能理性地同意受到某种方式的对待。既然同意原则谴责所有这一切行动，这个原则并未忽略实际同意的道德重要性。

192

人们反过来可能主张，这个原则暗暗地给予了实际同意以太大的重要性。考虑

否决权原则：以人们事实上或本会事实上拒绝同意的任何方式对待他们都是不当的。

正如类似的给予选择原则一样，这个原则也显然是错误的。有无数可允许的或道德上要求的行动是某些人要么事实上要么会拒绝同意的。例如，在地震情形中，即使格蕾拒绝同意，我也应当救怀特的命而不是格蕾的腿。通常也不存在任何可能的行动是每个人事实上会同意的。有人现在可能会论证：

以人们不能理性地同意的任何方式对待他们是不当的。

(J) 若人们事实上或本会事实上拒绝同意受到某种方式的对待，那么他们就不能理性地同意。

因此

以人们实际上拒绝或实际上会拒绝同意的任何方式对待他们都是不当的。

如果(J)是对的，同意原则就会蕴含否决权原则。这会使得同意原则显然是错误的。

我们应该接受(J)吗？询问人们是否能理性地同意他们实际上拒绝同意的行动，这可能令人困惑，因为这些人不可能同时既给出同意又拒绝同意。为了使得问题更为清楚，我们可诉诸另一版本的同意原则。按照

193 **CP3:** 如果人们若知道相关的事实则就不可能事先不可逆地同意对待他们的方式，那么这种方式就是不当的。

我们对于某个行动的同意是不可逆的，若我们知道，如果我们后来收回我们的同意，这对于我们以后受到何种对待毫无影响。

我们对许多行动都不能事先理性地给出这种不可逆的同意。例如，我们很少能事先理性地同意我们在行动当时实际拒绝同意的性交。因为绝大部分性活动的性质在很大程度受到在活动当时所有相关方是否实际上同意的影响，故这很少能是合理的。

然而，我们还是能对许多行动理性地给出这种不可逆的同意。我们要有充分的理由给出这种同意，下面的两种情况都得为真：

(K) 我们有某种理由给出不可逆的同意，由此限制了我们未来的自由。

并且

(L) 我们不会因在后来得知某种事实，从而会给予我们决定性的理由后悔我们早期给出的这种同意。

但这两个条件通常是能满足的。例如，在许多情形中，人们需要知道某人的同意是有约束力且不能收回的。设想在地震情形中，我一旦开始救怀特的命而不是格蕾的腿，再停止这样做就会很危险。接下来假定格蕾知道所有相关事实，在现在与将来有着同样好的决策能力。基于这些假设，格蕾现在能够理性地做出决策。理性并不要求我们尽可能地推迟决策。我讲过，格蕾会有充分的理由选择我救怀特的命而不是她的腿。如果情况果真如此，那么格蕾也有充分的理由对我后来那样做给出不可逆的同意。格蕾能理性地说，“去救怀特的命，即使我后来改变心意。”

当我们以CP3的形式应用同意原则时，我们的目标只是询问，人们是否能理性地同意以他们事实上拒绝同意的方式对待他们。我们若把它应用于事先给出的不可逆的同意，则这个问题更易回答。在许多实际情形中，人们事实上不会具有充分的理由，事先给出这种同意，由此使自己做出限制未来自由的某种承诺。但是给定我们虚构的思想试验的目标，我们可以假定，这些人会具有充分的理由事先做出他们的决策。我们的问题是，基于这个假定，这些人是否出否有充分的理由给出他们不可逆的同意。

194

我讲过，格蕾可以有这样的理由。我认为，人们在其他许多情形中能够理性地给出这种不可逆的同意——同意后来在无需其实际同意的情况下受到某种方式的对待。如果这为真，我们就能够拒绝上述论证中的前提(J)。同意原则并不蕴含否决权原则，至少避免了对那个原则最强有力的反驳。

尽管我们应当拒绝否决权原则，但我们可以言之成理地接受此原则的一个弱得多的版本。按照我们称为的

权利原则：在未经实际同意的情况下，每个人都有权不受到某种方式的对待。

当我们主张人们有权不受到某种方式的对待时，我们在部分程度上意指，未经这些人的同意，这种行动会是不当的。我们可称这些行动为涵盖否决权的行动。

在陈述这个原则时，我们通常很难判定哪些是人们有权否决的行动。这

种原则要想为人所接受，这些权利就必须得到严格描述。例如，我们不应该主张，每个人总是拥有不被杀的权利，因为有些杀害是无可避免的，有些杀害是有证成的，正如在自卫情形中那样。但是我们也可能主张，我们都有某种更受限制的权利，诸如未经我们的同意，即使为了我们自己的利益也不能杀掉我们。我们可能会同样主张，每个人对于如何处置其身体具有否决权，不仅仅是性方面的而且还有其他方面的。例如，基于一种观点，每个人都有权拒绝同意治疗而不再活着。

权利原则除了谴责当事人拒绝同意的那种涵盖否决权的行动外，还应该要求我们给予人们拒绝同意的机会。如果我们由于无法与人们交流而无法给予他们这种机会，那么我们应该尽力以涵盖否决权的方式（即他们如果有这样的机会，本会同意的某种方式）对待他们。当人们对某个行动没办法表示同意，但我们知道他们本会同意还是拒绝同意时，这些事实具有类似的道德重要性。当我们询问某人是否会事实上同意某个行动时，这完全不同于询问他们是否能理性地给出这种同意。我们也许知道某些人事实上不会同意某个涵盖否决权的行动，即使他们拒绝同意是不理性的。我们可能会说，在这样的情形中，人们有权不理性并承受其苦果。

然而，同意要想在道德上是重要的，给出同意的人必须对相关事实有充分的了解，并且能在头脑足够清醒的状态下考量这些事实。不过，满足这两个条件的人还是可能会做出不理性的决策。但权利原则不应该诉诸这样的同意：给出同意的人对最重要的相关事实并不了解，过于年轻，有严重的精神疾病，或受到其他扭曲性的影响（喝醉了、嗑药了和受威胁了）。可以说，在这样的条件下，人们无法有效地给出或拒绝同意。

若人们无法有效地对行动给出同意，则我们可能会问，如果这些人不受到这些扭曲性的影响，那么他们是否本会给出这样的同意。但这个问题可能很难回答。我们还有其他可行的方式来修正或拓展权利原则。与其诉诸我们认为在我们行动之时某人本会给出的假设同意，我们或许不如诉诸这个人早前所做的实际同意。在某些情形中，人们若知道他们随后会受到扭曲性的影响，可能事先就同意或拒绝同意随后受到某种方式的对待。我们可能认为，我们随后应该遵循这些早前有效的决策。还有些情形，人们当时无法给出有效的同意，并且他们事先也没有给出或拒绝同意。在这样的情形中，我们可能认为应当仅以如下方式对待他们：他们事后会追溯性认可，事后会很高兴我们的作为。“人们本会给出同意”这样的主张是无法确证的，“事后认可”这样的预测则不一样，有

许多是可以确证或证伪的。这对我们使用这样的预测来证成我们的行动能提供一个有效的检验。

我们接下来会限定权利原则，以便它能反映“有效同意的条件是个程度问题”这个事实。若人们的判断受到一定的扭曲性影响，但不足以使其决策无效，我们可以有证成地给予这种决策以较少的道德权重。

为了阐明其中某些要点，我们可以回到这个观点：在无法给出同意的情况下，每个人都有权不接受手术。人们通常主张这个权利是绝对的，毫无例外。但我认为，还是有些例外的。假定，在

手术情形中，为了救格林的命，我们在没有麻醉剂的情况下给他做手术。这个手术非常痛，但它会给格林带来多年有价值的生活。格林事先对这个手术给出了不可逆的同意，如果后来的痛苦导致他改变心意，那么在必要时就可以使用暴力。

在发现麻醉剂之前，许多人对救命手术都理性地给出了这种不可逆的同意。如果格林给出了这种同意，而后来的痛苦的确导致其改变了心意，那么我认为，我们可以有证成地使用武力来完成这个手术。权利原则应该允许这种行动。无论如何，我们会认为：既然剧痛是某种严重的扭曲因素，格林收回同意就是无效的。

我们接下来假定，在这种情形的另一个不同的版本中，格林拒绝事先给出这样的同意。我们可能认为，这个拒绝是决定性的，并推出我们应当让格林死去。但我们可能会反过来认为，格林的拒绝应该被看作是无效的，或得到的权重较少，因为即将到来的剧痛是另一个扭曲性的因素，使得人们难以做出合理的决策。基于权利原则的一种版本，如果手术的痛苦是短暂的，我们能有效成地强行给格林手术；并且我们还有很强的理由认为，格林事后会认可我们的决策，会很高兴我们救了她的命，尽管无论是事前还是当时她都拒绝同意。我们知道，在这样的情况下，最痛的时候一旦过去，大多数人就会认可这种手术。

197

然而，在这种情形中，还有另一种不那么明显的扭曲性因素。当我们考虑痛苦的经验时，我们大多数人都会对未来有一个强烈的偏向。一旦我们的痛苦结束，我们对它的关注就大为减少，甚至是毫不关注。由此，诉诸“一旦手术结束，人们几乎总是会回溯性认可这个行动”这一事实，就很难证成强制性地对做救人性命的手术。鉴于我们对未来的偏向，我们可能会低估“我们早前想

要避免而现在已经过去的那种痛苦”这种理由所具有的力度。

我们接下来假定，在

抑郁情形中，布鲁决定自杀。我们有强有力的理由认为，如果我们强制地阻止这种行动，布鲁的抑郁很快就会消失，余生都会过得很好。

我们许多人都会认为，我们能证成地否决布鲁的决策，使用武力阻止她自杀。如果我们接受权利原则，我们可能会主张：严重的抑郁是充分的扭曲性因素，由此布鲁的拒绝同意是无效的。但如果我们做出这个主张，我们的有效性标准将很高，且通常满足不了。严重抑郁的人可能知道相关的事实，也不是显然没有能力做出合理的决策。更可行的是这样的主张：尽管布鲁的抑郁并不会使得她的拒绝同意是无效的，但会使得她做出的决策不那么合理，因此她的拒绝在道德上可以被她其他时候的决策所否决。例如，如果布鲁有频繁的暂时性抑郁，她可能会事先同意，我们后面可以使用武力阻止她抑郁时的自杀。这就足以证成我们的行动，尽管在此我们的行动会驳回布鲁当时做出的有效的拒绝同意。由于自杀的不可逆性，这种行动甚至没有这种早期的同意也是可证成的。这里还有一个重要的不对称性。如果我们挫败了布鲁自杀的企图，她随后还能再次尝试；如果我们任其自杀，她后面就没机会活着了。

198

对于一种不同类型的例子，假定，在

错误信念情形中，我们能通过输血救布朗的命，但布朗拒绝同意，因为她是一名耶和华见证会教徒（Jehovah's Witness），认为输血是不当的。

我已经说过，人们要做出有效的同意，就必须知道相关的事实。布朗若知道相关的事实，就知道输血并不是不当的，于是就能理性地同意我们以这种方式救她的命。但我们也许会认为，既然布朗实际拒绝同意，我们以这种方式救她的命就是不当的。当人们因为具有某类错误的信念（比如某种道德或宗教信仰）而拒绝同意某个行动时，我们能可行地认为这种拒绝应该被看作是有效的。

我在这样说时假定了当前的同意比过去的同意更为重要，后者又比回溯性认可更为重要。值得询问的是，这种时间上的差别为何具有这样的重要性。

我若无法与你交流，就可能会尽力判别何种行动更有可能满足你的欲望和

偏好。正如我说过的，尽管自己的偏好只给予我们派生性理由，但我们具有非派生的理由来满足他人的偏好^①。我在努力满足你的偏好时，我没有任何理由对你现在的偏好给予优先性。假定我有理由认为：你现在想要我以某种方式行动，但你后来会改变心意，我若以另一种方式行动你会更高兴。我还有理由认为，你后面改变心意是因为你知道更多的相关事实，由此后面的偏好会有更根据。我认为，基于这些假设，我能理性地且有证成地优先满足你后面的偏好。

作为属于此类的一例，我们可以假定，作为你的医生，我必须决定是否以某种方式治疗你。由于你是昏迷的，我无法征得你的同意，只能尽力预测你的偏好与选择。这种治疗很快使你遭受某种痛苦，但使你以后能免于大得多的痛苦。我有很好的理由认为，你现在不喜欢我这样治疗你，但当你后来知道更大的痛苦是什么时，你会改变心意。鉴于这些事实，我可以有道理地认为，我应该满足你后来可预测的、更明智的偏好。

接下来假定在这种情形的不同版本中，你是有意识的，由此我可以就提议的治疗方案征询你的同意。如果你拒绝同意，这个事实可以清楚地否决我可行的预测：你后来会后悔做出这个决策。尽管我没有任何理由让你当前的偏好优先于后面的偏好，但你若有能力决定我如何治疗你，我确实有理由更为重视你现有的决定。

为了解释这种差别，我们首先可注意，我们对自己和他人信念在态度上存在着类似的事实。我若试图得到某个问题的真相并考虑他人的信念时，我就没有任何理由更为重视他人的当前信念。如果我有某种办法知道此人后面的信念，我就可能有良好的理由更为重视后面的信念，因为后面的信念有更好的根据。我还可能有理由更为重视别人早前的信念，因为它们较少受到扭曲性的影响。相对比而言，我必须更为重视自己的当前信念。例如，我能认为某个主张是错的，尽管我早前认为，或预测以后会认为它是对的。但不可能认为某个主张是错的，尽管我现在认为它是对的。我们永远都不可能把决策基于真理而不是我们现在相信为真的信念之上。

类似的主张也适用于我们的决策。我们必须更重视现在的判断，因为这些决断是基于我们现在相信为真的信念之上。即使我们的信念没发生变化，或者我们相信它们不会改变，我们也必须更为重视我们现在的判断，因为我们没法

^① 见原文第66页的讨论。

200 从过去或未来的视角来做出决策。我们必须从我们当前的视角来生活。这些事实能解释，当他人应该得到我们同意才能行动时，为什么他们应该更重视我们现在是否同意他们对待我们的方式。

28 道义式信念

同意原则声称只描述了我们行为不当的一种方式。行动即使能得到每个人的理性同意，依然可能是不当的。

许多这类行动是不当的，因为人们实际上不同意或实际上会不同意它们。正如我讲过的，这对于大部分直接干涉我们身体的行动而言，情况都是如此。另外更大一组的情形涉及所有权。人们对于我们如何处理他们的财产并不总是具有否决权，因为如果这是救人性命或使人免受伤害的唯一方式，那么即使所有者拒绝同意，我们仍然能有证成地使用甚至是破坏多种财产。但在许多情形中，未经个人的实际同意就使用或破坏其财产是不当的。如果我未经你的同意而住你的屋、穿你的衣和吃你的食物，那可能就是不当的。在大部分情形中，同意原则都会谴责这样的行动，因为我们不可能事先就理性地同意，其他人在当时不经同意就可这样行动。但如果我早前无家可归，饥寒交迫，那么这些事实可能会让你有充分的理由事先同意我这样做。同意原则由此就不会谴责我的行动。尽管有这个事实，若我未经你的实际同意，就住你的屋、穿你的衣、吃你的食物，这依然是不当的。

还有些行动，即使相关的每个人实际上且理性地给出了他们的有效同意，可能依然是不当的。例如，许多人对于自愿安乐死（如某人要求我们做的，为了这个人的自身利益而杀死他）都有这种看法。有些行动不当的理由不在于它们对待其他人的方式，因此就不会出现同意的问题。例如，残酷地对待动物就是这样，有些人相信自杀也是这样。

201 既然行动可能在其他方面或出于其他的理由而不当，同意原则的蕴含就可能部分地取决于哪些行动出于其他理由会是不当的。因此当我们应用这个原则时，我们有时必须诉诸我们有关“何种行动不当”的信念。我称这样的信念为道义式的（deontic），而行动的不当性所提供的理由可以称为道义式理由。

人们可能会反驳道，如果诉诸这些信念来应用同意原则，我们的道德推理会是循环的或者是乞题的。这种推理不能支持“何种行动不当”的信念。

这个反驳在部分程度上是正确的。下面两种不能同时为真

(M) 有些行动因为人们不能理性地同意而是不当的。

以及

(N) 这个人不能理性地同意它，是因为它是不当的。

某个行动若要因为某人不能理性地同意它而是不当的，则这个人必须具有决定性的非道义式理由来拒绝同意。但人们往往具有这种理由。例如，在地震情形中，怀特就有这样的理由拒绝同意我救格蕾的腿而不是他的命。怀特不能理性地同意这个行动，不是因为它是不当的，而是因为怀特死掉的损失比格蕾失去一条腿的损失要大得多。当我们把同意原则应用于这类和许多其他类情形时，它支持且有助于证成我们的道义式信念。

然而，正如我讲过的，我们有时候必须诉诸道义式信念来应用同意原则。假定在地震情形中的第二个版本中，我们可以称之为

手段情形，怀特和格雷两人被困于正逐渐坍塌的废墟中。尽管怀特的生命受到威胁，但格蕾毫无危险。我能救怀特的命，但只能在未经格蕾同意的情况下以其身体为盾牌，而这会毁坏她的腿。

我们许多人会认为，由于格蕾会拒绝同意，我以毁坏格蕾的腿的方式来救怀特就是不当的。基于这种（我们在此假定为真的）观点，若我们未经同意而严重伤害某人以有益于其他人，这种行动就是不当的。

202

我们把同意原则应用于这个情形时，首先可把“这个行动是不当的”这个假设放在一边。如果这个行动不是不当的，那么这个情形就与地震情形没有相关差别了。在地震情形与手段情形中，要么怀特死掉要么格蕾失去一条腿。这两个情形的差别只在于救怀特的命与格蕾失去腿之间的因果关联，而格蕾也没有强有力的理由要偏爱失去其腿的某种方式。我们还能假定，两种方式对她的损失都是一样的。我认为，在两种情形中，格蕾本来能理性地给出不可逆的同意，让我随后去救怀特的性命，即使格蕾会由此失去腿。并且在两种情形中，怀特的损失都比格蕾的要大得多，所以怀特都不可能理性地同意我不救她的命。基于这些假设，同意原则会要求，我在手段情形中以格蕾腿为代价救怀特

的命，因为这是怀特与格蕾都能理性地同意的唯一行动。

我们现在回到“这种行动是不当的”这个假设。如果同意原则要求这个不当的行动，这将是对此原则的强烈反驳。但我认为，该原则不会要求这种行动。如果我毁坏格蕾的腿救怀特是不当的，这种不当性就会给予怀特充分的理由同意我不这样做。我认为，如果有益于我们的方式会不当地伤害其他人，那么即使这种利益生命悠关，我们还是有充分的理由同意不这样做。

203 这里还有另一种方式捍卫这种信念。我们是在影响行动的意义上来讨论可能的同意。怀特要能给出同意或拒绝同意，我们就必须假定我已经给了怀特能力来选择我如何行动。如果怀特选择我损害格蕾的腿来救她的命，她就对我的不当行动负有部分责任。这就使得怀特做出这种选择是不当的。我认为，我们总是有充分的理由不做出道德上不当的选择。我这里并不是主张怀特做出这种选择是不理性的。也许怀特能理性地选择我行动不当，因为这个选择能救她的命，这就会给予她充分的理由做出这种选择。但是怀特反过来也有充分的理由不对这种不当行动负有部分责任。既然怀特能理性地同意我不以损害格蕾的腿来救她的命，同意原则就没有错误地要求这种不当的行动。

人们接下来可能反驳道，既然格蕾能理性地同意我以这种方式救怀特的命，同意原则就错误地允许了这种行动，甚至在这种行动由于格蕾实际上拒绝同意而不当的时候也是如此。但这个反驳误解了同意原则。这个原则宣称描述的只是使得行动不当的事实之一。因此，当这个原则并不谴责这种救怀特命的方式时，它并没有蕴含这种行动在道德上是允许的，从而允许这种行动。

这也适用于其他的情形。在我们正讨论的情形中，有些行动是不当的，但根本不是因为某人无法理性地同意这个行动，而是另有原因。我们可以论证：

仅当某人没有充分的理由同意不做某个行动时，同意原则才要求这样做。

(O) 只要某个行动出于其他理由而是不当的，这个行动的不当性就会给每个人以充分的理由同意我们不这样做。

因此

同意原则从不要“出于其他理由而不当”的行动。

我们可以类似地论证，这个原则从不谴责“出于其他理由而在道德上所要求”的行动。如果某个行动是被要求的，则所有其他的替代选择都是不当的，这将给予每个人以充分的理由同意这个行动。

基于某些观点，前提（O）是可以被否定的。假定，在

火灾情形中，布莱克陷于着火的废墟中，我若什么也不做，他马上就会缓慢且痛苦地死掉。我现在除了在不断升高的灼热迫使我逃离前杀死他之外，没其他办法让他免于这种痛苦。

接下来假定，布莱克知道这些事实而要求我杀死她。我认为，这个行动是能得到道德证成的。如果确实如此，布莱克就不可能理性地同意我不以这种无痛的较快死亡来有益于她。基于这些假设，同意原则要求我如布莱克所愿地杀死她。

基于某种观点，自愿的安乐死即使是在这种火灾情形中也是不当的。如果我让布莱克更好地死掉而有益于她是不当的，这个行动的不当性就会给布莱克充分的理由同意我不这样做吗？有些人回答不会。这些人可能同意，在手段情形中，怀特能理性地同意我不以损害格蕾腿的方式救她的命。但是怀特给出这种同意的理由，是由“我只能不当地伤害其他人来救她的命”这个事实所提供的。我们对于火灾情形并不能这样说。如果我应布莱克之请杀死她，我不会不当地伤害任何其他人。这些人可能会认为，既然有这种差别，我杀害布莱克的不当性就不会给予她充分的理由同意我不这样有益于她。基于这些假设，前提（O）在此会是错误的，同意原则将会要求一个不当的行动。

我认为，这个例子并没有对同意原则提出强有力的反驳。极少有人会同时相信：这个行动是不当的，且其不当性不会给予布莱克充分的理由同意我不这样做。我们拒绝这种观点是说得通的。

我们接下来考虑另一版本的火灾情形。假定，尽管布莱克知道我杀死她对她是有益的，她拒绝同意。有些人会相信这两点：没有布莱克的同意，这个行动是不当的；布莱克本不可能事先理性地同意：在未经她后来同意的情况下，我就不给予她这种更好而快的死亡。如果这两个信念都是对的，前提（O）就是错误的，因为同意原则在此会要求我行动不当。但我认为，如果我当时未经布莱克的实际同意而杀死她是不当的，那么这种行动的不当性也同样会给予布莱克以充分的理由事先同意我不这样做。

考虑一个不同类型的例子，假定，在

父母情形中，在船舶遇难后，你我各有一个孩子有生命危险。我有一个救生圈，能救你的孩子或我的孩子。

接下来假定，正如大多数人会相信的，我应当救我的孩子。你能理性地同意我这样做吗？

基于一种观点，答案是不能。如果我给你能力选择我要如何行动，你应当选择我行动不当来救你的孩子。尽管你对我的行动不当负有部分责任，但你保护自己孩子的义务在道德上会压倒你不选择我行动不当的理由。既然有这个事实，且你还有想要我救你孩子的其他强有力的理由，你就不可能理性地同意我不这样做。基于这些假设，同意原则在此会要求我行动不当，救你的孩子而不是我的孩子。

如果我们接受这种观点，且我们对其他在相关方面类似的情形具有类似的信念，则我们不得不修正同意原则，由此不再适用于这类情形。按照

CP4：以人们没有充分理由同意的任何方式对待他们都是不当的，除非这些人没有这种理由是出于这样的例外情况：涉及相冲突的、个人相对的道德义务。

206 这个修正尽管会限制同意原则的范围，但不会降低这个原则的可行性。当我们应用这个原则时，我们可诉诸一种思想试验，询问其他人是否能够理性地选择我们的行动方式。选择如何行动的人是我们还是其他人若会产生道德影响，那么我们那样询问就是没用的。在此类情形中，我们去做其他人选择我们去做行动就是不当的。我们这里的思想试验会引导我们忽略这种事实。我们不应该指望，在这样的情形中，同意原则能够帮助我们决定何种行动是不当的。既然我们能够解释，这个原则为什么不该应用于这类情形，那么这种情形就无法质疑这个原则所表达的道德观念。

然而，这个修正可能是没必要的。我们可以问

Q1：我们会有义务选择或使得他人行动不当吗？

基于某些道德观点，有时候答案是有。这样的一种观点是某类道德国家主义，它在第一次世界大战之前及一战期间的欧洲被广为接受。基于这种观点，如果你的国家与我的国家正在开战，那么我的爱国义务就是让你行动不当，泄露你国的军事机密，从而可以让我军击败你军。

康德对Q1的答案是没有。如果我们接受这个答案是正确的，那么父母情形并没有破坏这个康德式理想。基于这样一种观点，我们能具有在某种意义上是相冲突的、个人相对的义务。情况可能是我有救我的孩子的义务，你也有救你的孩子的义务，尽管我履行自己的义务会让你无法履行你的义务。但是在父母情形中，我能以你能理性地同意的方式行动。既然我救你的孩子而不是自己的孩子是不当的，你就不可能具有义务选择我这样做，这种行动的不当性会给你充分的理由同意我尽自己的义务，救我的孩子。

作为另一种反驳，接下来假定，在

平等要求权情形中，我能救你的命或格蕾的命。

在这种情形中，看来你不能理性地同意我救格蕾的命而不是你的命。你有很强的个人理由不给出这样的同意。既然你的死亡不偏不倚地与格蕾的一样坏，这些个人理由看来就是决定性的。格蕾会有同样的理由不同意我救你的命而不是她的命。同意原则看起来在此是失败的，由于不管我做什么，我都会以某人无法理性地同意的方式对待他，因此它错误地蕴含了我总是会行动不当。然而，我们能够可行地主张，我应当给予你和格蕾同等的得救机会。如果我不给予你和格蕾同等的机会是不当的，那么这个事实会给予你们充分的理由同意这个行动。

207

对于另一个例子，假定，在

高价格情形中，我卖给你一些产品，你本来能在其他地方非常便宜地买得到——产品的这一情况只有我知道。

接下来假定，由于你不是很富裕，你本不能理性地选择支付这种更高的价格。于是同意原则蕴含着，我在赚你的钱时行动不当。我们有些人会认为，既然你自由地同意了我赚你的钱，我并没有行动不当。但是同意原则在此并不是显然错误的。我们能够可行地认为，正如若这个产品有缺陷我应当提醒你一

样，我也应当告知你，你能在其他地方更便宜地买到这个产品。

我对这些情形的评论并不证明，我们总是能有证成地遵循同意原则，由此实现康德的理想。有些人会拒绝这些主张。还可能存在着其他的情形，其中没有任何可能的行动是每个人都能理性地同意的。但是，正如各种各样的情形所表明的，同意原则具有的含义通常是可行的，且绝不会陷入明显的错误。这使得值得询问：在道德和合理性的最可行的观点之中，何种观点是与康德的理想相容的。

29 极端要求

然而，对这个原则还有另一个反驳。假定，在

自我情形中，我与怀特被困于正逐渐坍塌的废墟中，我能够救怀特的命或自己的腿。

208 基于某些观点，在道德上这个情形与地震情形是一样的。我应当救怀特的命而不是自己的腿，因为怀特的损失会比我的大得多。我们大多数人都有不同的看法。基于这种看法，尽管我挽救某个陌生人的腿而不是怀特的生命会是不当的，但在道德上允许我救自己的腿。我们在代价很小的情况下应当救陌生人的腿，但这里我要付出的代价太大。

同意原则在此蕴含着什么呢？如果怀特能选择我如何行动，他能理性地选择我挽救自己的腿而不是她的生命呢？

答案看样子是不能。情况看来是，怀特不可能理性地同意任何人挽救自己的腿而不是她的命。但这种观点太过简单。我们有理由不仅关注会发生什么，还要关注谁会做这些事情，为什么会做这些事情。

为了阐明这一点，降低筹码会是有帮助的。首先假定我能使怀特免于了一星期的痛苦，或者使某个陌生人免于一天同样痛苦。怀特和这个陌生人之间并没有任何其他的相关差别。基于那个假设，我没有任何理由不那么看重怀特的福祉。怀特不可能理性地同意我的这一选择，即毫无理由地选择帮那个陌生人而不是帮她减除更大的负担。这个选择对怀特的对待好象她是低劣者，或者当她根本不存在。

我们接下来假定，我要么让怀特免于了一星期的痛苦，要么使自己免于一天

的痛苦。尽管我没有理由更为关注某个陌生人的痛苦，但确实有理由更为关注自己的福祉。我们都有理由更为关注发生在自己身上的事情。既然每个人都有这样的理由，我们就能理性地同意，出于这样的理由，其他人能更为重视自己的福祉。尽管怀特不能同意，我毫无根据地选择不是使她免于了一星期的痛苦，而是使陌生人免于了一天的痛苦，但怀特具有充分的理由同意我可以让自己免于这一天的这种痛苦。对待怀特的这种方式没有把她好象当作一个低劣者，或者当她根本不存在。

然而，在自我情形中，筹码要高得多。怀特可能没有充分的理由同意我救自己的腿而不是她的命。

如果正如大部分人所认为的，我在道德上允许救自己的腿而不是怀特的命，这会产生影响吗？也许不会。在此，可允许性与不当性之间可能存在差别。如果我只能以不当行动来救怀特的命，那么我讲过，正如我们在手段情形中所认定的那样，这个行动的不当性会给予怀特充分的理由同意我不救她的命。然而，在自我情形中，我能够救怀特的命而不用行动不当。即使道德允许我救自己的腿而不是怀特的命，这个行动的可允许性也不会给予怀特充分的理由同意我不救她的命。

这个行动的可允许性若不给予怀特这种理由，则她就不能理性地同意我不救她的命，这样同意原则会要求我救她的命而不是自己的腿。这个原则在此会与多数人的信念相冲突。

若只能自己付出巨大代价才能救他人的命，那么这样做的人会很少，类似的推理还适用于许多情形。我们许多人往往可以要么有益于自己要么给予他人更大的利益。当给他人的利益要大得多时，这些人可能没有充分的理由不有益于他们。假定，在

援助机构情形，我能够花200美金去夜场娱乐，或者是把钱给予高效的援助机构，诸如Oxfam，它可以用这个钱使得遥远之地的穷人免于死亡、不至于瞎掉或者免于其他巨大的伤害。

同意原则若应用于这两个选项，看来就蕴含着我们应当把钱捐给援助机构。穷人看来没有充分的理由同意我不这样做。这在明天乃至每一天都适用于我。这在每一天也同样适用于本书的大多数读者。有超过十亿的人每天只有2美金左右，与他们相比，本书的绝大多数读者都是非常富裕的。

如果出于这些理由，同意原则要求富裕者把许多财富或收入转移给穷人，这对此原则根本就不是反驳。既然富裕者能够很容易地让许多穷人免于死亡或痛苦，任何讲得通的原则或道德理论就都会做出同样强烈的要求。富裕者尽管在法律上有权拥有其一切财产，但在道德上有权拥有的财产要少得多。康德写道：

拥有财力去实践诸如行善之类依赖于命运眷顾的德行，这大多是某些人通过……不义而受惠的结果。

据说他还讲过

一个人可能参与一般的不义，纵然他从不行不义之举……甚至慷慨之举也是尽责和感恩之举，这是他人的权利所引出的。

然而，同意原则可能要求太高了。在严肃地思考正义要求和考虑相关论证之后，我们可能不得不承认，我们富裕者全部应当把财富或收入的十分之一、五分之一或三分之一转移到第三世界。但是同意原则所要求的远不止这一点。这个原则若要求过高则可修正。我们可能会主张：

CP5: 以人们没有充分理由同意的方式对待他们是不当的，只有当为了避免这种行动我们会承担太大的负担时才能例外。

在应用这个版本的同意原则时，我们不得不判定这个负担在什么时候才是太大的。当我们考虑极端的全球不平等所引起的道德问题时，这是一个极困难的问题。一个问题是我們是否应该以及如何来评估许多小馈赠的累积代价。但我们起始可以这样主张：在自我情形中，我被允许救自己的腿而不是怀特的命。

如果同意原则要求过高，且必须在在这方面予以弱化，康德的人际关系理想看来在原则上就是不可能的，因为存在着某些正当行动是有些人无法理性地同意的。但这些行动只在它们在道德上被允许这个意义上是正当的。道德上要求的行动，可能没有任何一个是人们无法理性地同意的。这样我们就仍然能实现康德的理想。每个人只以每个人都能理性地同意的方式行动就依然是可能的。也许总是至少存在一种行动是正当的。例如，在自我情形中，我可以救怀

特的命而不是自己的腿，这个可敬的行动会是正当的。如果同意原则要求太高，这最多是蕴含着：为了实现康德的理想，我们可能为彼此所做的不得不超出道德要求我们所做的事情。这一点也不让人吃惊。

我的结论是，我们有很强的理由接受某个版本的同意原则。这个原则可能要求太高，也可能存在着修正它的其他方式。但是至少在大部分情形中，以人们不能理性地同意的方式对待他们是不当的。若我们的行动影响很多人，并且只有一个行动是每个人都能理性地同意的，这个事实会给我们很强的理由这样做，并且足以解释道德为什么会要求这样做。按照某些可行的假设，同意原则永远不会偏离轨道，从而要求出于其他理由而不当的行动，或者谴责被要求的行动。

然而，同意原则不可能是康德努力寻找的那个东西：道德的最高原则。即使每个人能理性地同意某些行动，它们还可能是不当的。同意原则陈述了康德的人性公式所表达的观念之一。既然我们至少还需要一个原则，我们现在就可转向公式的这个部分。

第九章 纯粹作为手段

30 纯粹手段原则

212 通常有人主张，使用人是不当的。但这个主张需要限定。如果我们一起爬山，我可能会站在你的肩膀上，把你用作梯子。我可能会问你某个字的意思，把你用作字典，或者让你见证我遗嘱的签名，把你用作证人。这种用人的方式并不是不当的。康德主张，不当之处在于纯粹使用人们。正如其他人所说的，“你只是利用我”。

按照我们所称为的

纯粹手段原则：把人们纯粹作为手段来对待是不当的。

不纯粹地使用人们，我们能如何使用人们呢？在解释这个区分时，我们首先比较两个科学家如何对待实验室中的动物。人们可以假定，一个科学家以最有成效的方式做试验，不管这给动物带来多大的痛苦。这个科学把动物纯粹作为手段来对待。另一个科学家只以不给动物带来痛苦的方式做试验，尽管她知道这些方法不会那么有成效。这个科学家像第一个一样，把动物作为手段，但她没有把它纯粹作为手段，因为她对它们的使用受制于她对其福祉的关切。

这也同样适用于我们彼此的对待。按照一种粗略的定义

213 若我们利用一个人的能力、活动或身体来帮助我们实现某个目的，我们就是把他们作为手段来对待。

这个定义需要以某种方式来加以限定。例如，我们有时候应该区分：对某人做某事作为实现某个目标的手段和把这个人作为手段来对待。例如，医生为了发现我肋骨是否断裂，他压迫我的整个胸部，说，“告诉我哪里疼”。这里医生是在使用我的身体，弄疼我，作为发现这个信息的手段。但他没有把我作为手段来对待。为了涵盖这种情形，我们可以提出，我们的目标若是为了某人好，且我们的行动得到了这个人的同意，则我们并没有把他作为手段对待。

按照另一种粗略的定义

如果我把某个人作为一种手段对待，且视此人为纯粹的工具，我们就是把这个人纯粹作为手段来对待：我们忽略这个人的福祉和道德诉求，我们会以任何能最好实现我们目标的方式来对待他。

卡姆拒绝了这个定义。她驳斥道，如果基于康德的原则，我们永远不应把人们纯粹作为手段对待的意思就是如此，那么这个原则就太弱，也太容易遵循了。例如，基于这个定义，某个奴隶主若对于其奴隶的福祉略有关照，不让奴隶在天气最热时工作，则他就没有将其奴隶纯粹作为手段来对待。但这个人对待奴隶的方式肯定会受到康德原则的谴责。

我认为，这个反驳表明我们不仅应当修正这个定义，还应当修正康德的原则。

对于一个类似的例子，考虑康德的例子

(A) 富裕者不给穷人任何东西是不当的。

假定某个富人终其一生总共给了穷人1美金3美分。既然这个人给了穷人一点东西，(A) 并不蕴含他行动不当。正如这个例子表明的，(A) 太弱，因为这个人没有给予更多是不当的。我们应该主张，如果富裕者给予穷人的太少，他们也是行动不当。这类不当性是个程度上的问题。 214

我们可能会主张，把人们纯粹作为手段来对待的不当性也是如此。基于一种更强形式的康德原则，我们可称之为

第二种纯粹手段原则：把任何人纯粹作为手段对待或接近如此都是不当的。

若我们把某人作为手段对待，且给予此人的福祉或道德主张的权重太低，我们就是接近把他纯粹作为手段来对待。这就是我所虚构的奴隶主对待其奴隶的方式，即使他让其奴隶在最热时休息。于是这个修正的原则也谴责这个奴隶主的行动。

我们接下来能主张

(B) 我们并没有把某人纯粹作为手段对待，或接近如此，如果
要么

(1) 我们对待某人的方式充分重要的意义上受到了相关的道德
信念或关注的管辖或指导。

要么

(2) 我们确实或会在相关方面选择为了此人的缘故而承担巨大的
负担。

215 某个道德信念要想在(1)所意图的意义上是相关的，就必须要求我们直接关注以何种方式对待那个人的福祉或道德主张。假定另一个奴隶主从不鞭打其奴隶，因为他认为这种行动是不当的。但他认为，使得此行动不当的不是这会给其奴隶带来痛苦这个事实，而是他会从中得到虐待式快感这个事实。如果这就是他从不鞭打其奴隶的缘由，这个事实无法算作反对这一指控：他将其奴隶纯粹作为手段对待。另一个例子是康德认为，残酷地对待动物之所以不当，是因为它蒙蔽了我们的同情心，使得我们更有可能残酷地对待其他人。某个科学家如果只是由于这种道德信念才避免给其实验动物带来痛苦，则还是在将动物纯粹作为手段来对待。

既然相关性和重要性都是程度上的问题，(1)是否为真通常就是不清楚的。有些奴隶主之所以不鞭笞其奴隶，是因为他们关注其奴隶的福祉。这个关注尽管是相关的，但没有以足够重要的方式管辖这个人的行动。下面是个不那么清晰的情形：当我的母亲1930年代在中国的一条河上旅行时，她的小船受到一帮土匪的劫持，他们的道德原则只允许他们从普通人那里拿走一半的财产。这些土匪允许我妈选择，让他们拿走订婚戒指还是结婚戒指。如果说这些人是把母亲作为手段对待，那么他们并没有把她纯粹作为手段来对待。他们近乎

那样做吗？我倾向于回答没有。但这可能是个边缘性情形，其中我们对这个问题没有确定的答案。

要满足条件（2），我们只准备为他人的缘故承受巨大的负担是不够的。这个事实可能与我们所考虑的行动不是充分相关。考虑某个爱妻的人，他在某种灾难中宁愿为了她而付出自己的生命。如下情况仍然是可能的：在这个人的大多数日常家庭生活中，他把她纯粹作为手段来对待。

我们是否把某人作为手段来对待，只取决于我们的意图。我认为，我们是否把某人纯粹作为手段来对待，还取决于我们根本的态度或策略。而这在部分程度上就是这样的问题：如果情况有所不同，那么我们会做什么。我们现在回到在研究中使用实验动物的那两个科学家。我们假定，在一次实验中，两个科学家都用最有效的方法，且不会给动物带来痛苦。尽管这两个科学家的行动方式一样，但第一个科学家仍是把动物纯粹作为手段来对待，因为情况仍然会是：即使最有效的方法会给其动物带来巨大的痛苦，她还是会用这种方法。第二个科学家没有将其动物纯粹作为手段来对待，因为她不会那样做。我们接下来考虑这样的主张：

216

他把她纯粹作为手段来对待。

在这个场合，在其所作所为中，他把她纯粹作为手段来对待。

第一个主张是更为自然的，这种主张是否为真通常也是更清楚的。

康德主张，把任何理性存在者纯粹作为手段来对待是不当的。基于类似但更宽泛的观点，把任何有感受性或有意识的存在者纯粹作为手段来对待是不当的。这些观点正确地蕴含着：把任何理性或感受性存在者视作纯粹的工具，由此我们可以随心所欲地对待它们，这是不当的。但是康德的主张看来还蕴含着，我们若把任何人纯粹作为手段对待就是行动不当。

这可能并不为真。考虑某个暴徒，他不像我母亲碰到的有原则的土匪那样，而是把大部分其他人纯粹作为手段，一旦伤害他们能有益于自己，他就会这么做。当这个人买咖啡时，他对待卖咖啡者如同对待自动售卖机。只要认为值得费神，他就会偷窃卖咖啡者的咖啡，正如他会为此砸碎售卖机。尽管这个暴徒把这个卖咖啡者纯粹作为手段对待，但不当的是他对这个卖咖啡者的态度。在买咖啡时，他并没有行动不当。

我们接下来考虑利己主义者，他以自认最利己的任何方式对待别人。康德

主张：

一个人若想许假诺……就是想把另一个人纯粹作为手段加以利用。

我们可以同样主张，当利己主义者由于需要许诺对象以后提供帮助而信守其诺言时，他就是想要利用这个人，把他纯粹作为手段来对待。我们接下来假定，利己主义者冒着巨大的危险救了某个溺水的孩子，但他唯一的目标就是获得报偿。康德的原则蕴含着，既然这个人把别人纯粹作为手段来对待，这个人在守诺和救孩子时是行动不当的。这显然是错误的。

为了避免这类结论，我们也许主张：

217

(3) 如我们所知，如果我们的行动不会伤害某人，我们就没有把这人纯粹作为手段来对待。

但是假定，在

互利的情形中，格林嫁给了90岁的亿万富翁，苟尔德。格林对苟尔德服务周到，在各面对他很好。正如苟尔德所知，格林的唯一目标就是继承他的部分财富。苟尔德尽管宁愿得到格林的真情，还是基于格林的利己条款接受了互利的安排。

接下来假定，格林视苟尔德为纯粹的工具，她以实现自己目标的最好方式对待他。格林的最初方案是仿造遗嘱，然后谋杀苟尔德；随后她改变了计划，嫁给了他，待他很好，仅仅是因为这是得到苟尔德部分财富的稳妥方法。按照(3)来说，既然格林知道她的行动不会伤害苟尔德，她就没有把苟尔德纯粹作为手段对待。这个主张是说不通的。尽管格林知道她的行动不会伤害格林，但这个事实对她的决策没有任何影响。如果谋杀苟尔德是个稳妥的办法，她就会那样做。我认为，我们应该承认，格林是把苟尔德纯粹作为手段来对待。

如果我们不能诉诸(3)，康德的观点就蕴含着格林行动不当。也许我们应该接受那个结论。但是当利己主义者信守诺言、冒生命危险救溺水儿童时，我们不应该主张这些行动是不当的。我们的主张应该只是，鉴于其自利的动机，这个人的行动就没有康德所称的道德价值。

为了避免谴责这种行动，我们也许会再次修正康德的观点，按照

第三种纯粹手段原则：把任何人纯粹作为手段来对待，或接近如此做，且这种做法还可能伤害这个人，我们的做法就是不当的。

在转向这个原则时，我们是在放弃这种观点：如果我们把某人纯粹作为手段来对待或接近如此，这就足以使得我们的行动不当。

我们已经讨论过，基于康德的观点，我们应当如何对待一切理性存在者或人的两种方式。第一，我们应当遵循同意原则，以人们能够理性地同意的方式去对待他们。第二，以任何方式把人纯粹作为手段来对待都是不当的。基于我们对第二个主张最新的版本，这样的行动仅当很可能会伤害那人时才是不当的。

218

我们接下来可以把康德观点的这两个部分联结起来。若我们对某人的对待在充分重要的意义上受到相关道德信念或原则的管辖或指导，那么我们就没有把他纯粹作为手段来对待，甚至也没接近于如此。康德自己的例子是同意原则。康德主张，如果我们有意只以人们能够理性地同意的方式对待他们，我们就把他们作为目的，而不是纯粹作为手段来对待。

我们现在回到

救生艇情形：我搁浅于一块岩石上，另五个人搁浅于另一块岩石上。在上升的潮水淹死我们所有人之前，你能用一个救生艇救我或救另外五个人。

我们还考虑

隧道情形：一辆无人驾驶的失控列车正冲向一个隧道，会杀死相同的那五个人。作为一个旁观者，你改变轨道，使得列车转向另一个轨道，穿过另一条隧道，从而能救这五个人的命。正如你所知，不幸的是，我处在这个隧道之中。

桥梁情形：列车正冲向这五个人，但没有其他的轨道与隧道。我在轨道之上的桥梁上。你救这五个人的唯一方法就是通过遥控打开我正站着的那块桥板活门，这样我会掉在火车前，因此激发列车的自动刹车。

在这三种情形中，如果你救那五个人，我都会死。但是我的死亡与你救那五个人有着不同的因果关联。在救生艇情形中，你会让我死是因为在可用的时间内，你无法既救我又救那五个人。在隧道情形中，你会使列车转向救那五个人，而可预见的副效果就是因此杀死我。在桥梁情形中，你把杀死我作为手段来救那五个人。我们能够假定，我和这五个人都是差不多的年纪，没人对我们的生命威胁负有责任，在我们之间也没有任何其他相关的道德差别。

人们也许会主张，在桥梁情形中，你并不是真正地把杀我作为手段来救五个人。你纯粹是用我的身体作为工具来阻止列车，如果我活下来你会很高兴。基于这种观点，仅当一个人的死亡是我实现自己目标的不可或缺的部分时，我杀他才是作为手段。例如，中世纪国王的第二个儿子想要成为王位的合法继承人时，情况就可能是如此：只有他的哥哥死掉才能实现这个目标。然而，在一种更宽泛的意义上，当我们以某种牵涉并可预见地杀死或伤害某人的方式行动以作为实现某个目标的手段时，我们就是把杀死或伤害某人作为手段。我就是这样使用“把杀死或伤害作为一种手段”这个术语的。

在救生艇情形中，大多数人认为，你要么可以要么应当救五个人的命。在隧道情形和桥梁情形中，有些人会认为你救五个人是不当的。基于这种观点，我们不杀人的义务要压倒或者优先于我们救人的义务。还有许多人认为，尽管我们不杀人的义务确实具有这种优先性，但是在隧道情形中情况不是这样的。基于这些人的观点，转变一些无意图且有威胁性的过程（诸如洪水、雪崩或者正失控的列车），这样它会杀死较少的人，这并非是不当的。在持这种观点的人中间，大部分会认为，如果在桥梁情形中，你杀死我作为一种手段来阻止电车以救五个人，你的行动会不当。也有些人拒绝这种区分，认为在所有这些类型的情形中，我们应当救尽可能多的人。我在此的目标不是要解决这个分歧，而只是询问我们一直在讨论的康德式原则蕴含什么。

我讲过，在救生艇情形中，我能理性地同意你救那五个人而不是我。如果这个选择是我的，我就有充分的理由救自己的命，但我也充分的理由救那五个人而不是自己的命。既然我还能理性地同意你救五个人，同意原则就不会谴责这个行动。

这也同样适用于隧道情形。如以前一样，如果选择是自己的，我就会有充分的理由救自己或是那五个人。在此我为了救那五个人而使列车转向，乃至反而杀死我——这不会引出任何相关的差别。我们可以设想，这种死亡方式对我来说绝不算坏。既然我能理性地使列车转向来救那五个人，自然也能理性地同意你这样做。因此，同意原则不会谴责这个行动。

这也同样适用于桥梁情形，其中你只能杀了我救这五个人。如果选择是我的，我会有充分的理由跳到列车前，这样它会杀死我而不是那五个人。相对于在隧道情形中杀死我是作为救五个人的副效果，在桥梁情形中，杀死我是救五个人的手段，这对我来说绝不会比隧道情形更糟。既然我能理性地自杀作为救五个人的手段，自然也能理性地同意你这样来对待我。

人们也许反驳道，我不能理性地同意你把杀死我作为手段，因为这个行动会是不当的。但如果我同意这个行动，它就不是不当的。这样，即使这个行动在未经我同意的情况下是不当的，这也不会给予我任何理由拒绝同意。

我们接下来假定，正如我所知，你接受同意原则并一直践行它，这样这个原则管辖着你的行动。如果有时间，我可能就会如此思考：

按照这个原则来说，以人们不能理性地同意的方式对待他们是不当的。
我能理性地同意你杀死我作为救那五个人的手段。

因此

即使我事实上不会同意，同意原则也不会谴责这个行动。

如果我们对待人们的方式受到同意原则的管辖，我们就没有把他们纯粹作为手段来对待。

221

因此

既然你对我的对待受到同意原则的管辖，你就既没有把我纯粹作为手段来对待，也没有接近如此，这样，没有任何版本的纯粹手段原则会谴责你的行动。

我认为，这个论证是健全的。未经我的同意，你杀死我作为救那五个人的手段也许是不当的。但康德式原则并没有蕴含这一点。

31 作为手段与纯粹作为手段

在做出这些主张时，情况似乎是我必定误解或误用了纯粹手段原则。基于

一种广泛接受的观点，我将称为

标准观点，如果我们在未经他人同意的情况下伤害他们以作为实现某个目标的手段，那我们由此就把这些人纯粹作为手段来对待，这种方式使得我们的行动是不当的。

我认为，这种观点涉及三个错误。第一，当我们把伤害人们作为手段时，我们可能没有把这些人作为手段来对待。第二，即使我们是在把这些人作为手段来对待，我们也可能没有把他们纯粹作为手段来对待。第三，即使我们把这些纯粹作为手段来对待，我们也可能没有行动不当。

我们首先假定，在

自卫情形中，当布朗以刀攻击而试图杀死我时，我踢他救自己，且预知会踢断他的腿。

222 尽管我把伤害布朗作为一种手段来阻止他杀害我，但我没有把他作为手段来对待。正如我们穿雨衣使自己不变成落汤鸡时没有利用降雨一样，当保护自己不受别人的攻击时我们也没有利用他们。我们能够补充道，我尽管应当把布朗本人作为目的而不是纯粹作为手段来对待，但我应该把伤害布朗纯粹作为手段，即使部分地作为目的或部分地为了其自身的缘故都是不应当的。

人们可能会反驳道，既然伤害某人是对待这个人的方式，把伤害某人作为手段就必定是把这人作为手段对待的方式。但这个反驳忽略了把“对某人做某事”作为手段与“利用这个人”之间的区分。正如我说过的，当我的医生伤害我以判断我的肋骨是否断了时，他并没有因此在利用我。我认为，他没有把我作为一种手段来对待，因为他是经我同意且是为了我自己的利益而伤害我。尽管我阻止布朗谋杀可能是有益于他的，但这并不是最好的方式解释：我为什么把伤害布朗作为手段时没有利用他。我们可能反过来认为，既然我纯粹是保护自己不受布朗的攻击，那么如果布朗不在那里，我的目标就更容易实现。如果我是在利用布朗，我就会想要他在那儿。

我们接下来转向这样的情形，其中我们在把伤害人们作为手段之时也确实把这些人作为手段来对待。基于标准观点，如果我们强行给某人带来伤害以作为实现某个目标的手段，这就足以使得我们是在把这个人纯粹作为手段来对

待。为了检验这种观点，我们考虑

第三种地震：你和孩子陷在正坍塌的废墟之中，这威胁到了你们两人的生命。你要救孩子，就得不经布莱克的同意把他用作盾牌，粉碎他一个脚趾头。如果你使他再碎掉一个脚趾头，你还能救自己的命。

假定你认为，你用这样的方式救自己的命是不当的。你认为，只有救孩子的命才能证成给别人施加这种伤害。按照这种信念，你救了孩子，布莱克因此只失去了一个脚趾头。由于你的行动未经布莱克的同意而伤害了他，标准的观点蕴含着你把布莱克纯粹作为了手段。但这不是真的。如果你把布莱克纯粹作为手段来对待，你就会既救孩子的命也会救自己的命，使得布莱克失去两个脚趾头。我们如果宁死也不愿给别人带来较小的伤害，就没有把别人纯粹作为手段来对待。

223

标准观点可以得到修正。有人可能表示，尽管你没有把布莱克作为纯粹的手段，但那是因为你限制了你带给布莱克的伤害，这种方式对你来说更糟，或者说没有那么有效地实现你的目标。在桥梁情形中，如果你杀了我救那五个人，对你的行动就不能这样说。你没有限制你带给我的伤害。即使你把我视作一个纯粹的手段，你还是会做完全相同的行动。这看来足够证成这个指控：在桥梁情形中，你这样的行动就是把我纯粹当作手段。基于这种建议：

(C) 我们把某人纯粹当作手段来对待，如果

(1) 我们在未经这个人同意的情况下将伤害他作为实现某个目标的手段。

除非

(2) 我们限制所带来的这种伤害，且这会或可能会严重不利于我们，或者使得我们的行动在实现自己的目标上远没有那么有效。

我认为这种观点也是错误的。我们已经假定，在第三种地震情形中，你决定不以牺牲布莱克的一个脚趾头而救自己的命。我们接下来假定，就在你行动之前，情况有了变化，这个正坍塌的废墟现在只威胁你孩子的生命。当你让布

莱克失去一个脚趾头时，你没有限制你带给布莱克的损害，这样（C）蕴含着
你正把布莱克纯粹当作手段。这是一个得不到证成的结论。你宁死也不想再使
布莱克失去一个脚趾头，这就足以表明你没有把布莱克纯粹当作手段。你现在
无法这样做是完全不相关的。

再举另一个例子，假定我是某场正义之战中的战士，与我的军队一起攻打
一个被占领的城市。在攻击任何建筑物里的敌军之前，我冒着生命危险，顶着
对方的枪林弹雨向他们喊话，给予他们投降的机会。如果这些人拒绝这样的机
会，而我以击杀或伤害他们为手段来占领某个建筑，那么（C）正确地认可，
既然我为了这些人的缘故而冒了生命危险，我就没有把他们纯粹当作手段。我
接下来假定，某建筑物中的敌军已经得到了投降的机会，但他们拒绝同意。
按照（C）来说，如果我杀害或伤害了他们，我就是把他们纯粹当作手段。这
并不为真，因为我本会冒着生命危险给他们投降的机会。在这个场合下，这
些人已经有了这样的机会，所以我才没有这样做，但这是不相关的。我对于所有
敌军的态度是一样的，我没把他们中的任何人当作纯粹的手段。

这也适用于桥梁情形。假定你用遥控使我掉在轨道上，这样我的身体会阻
止正在奔驰的列车。你的目标是确保那五个人得救。然而，你也跑向轨道，试
图跳在列车之前，从而在列车撞上我之前阻止它，救我的命。如果你的企图成
功了，你就没有把我当作纯粹的手段，因为你能为了我的缘故牺牲自己。我认
为，即使你没有及时赶到，这也不会产生任何分别。如果你本来宁愿牺牲自己
以免杀死我，即使这永远都没有可能，这也不会产生任何分别。在桥梁情形
的两个版本中，你的行动都可能是不当的。如果这个行动是不当的，使其不当
的是“你会杀死我作为手段来救那五个人”这个事实。但是你没有把我当作纯
粹的手段来对待。

我已经拒绝了有关把人们作为纯粹的手段来对待的标准解释。某些作者给
出了其他的解释。例如，奥尼尔写道：

如果我们强制或欺骗他人……就确实利用了他们，在我们的计划
中把他们当作了纯粹的手段或工具……欺骗或强制的准则就把其他人
当作了纯粹的手段……

科丝嘉同样写道：

强制和欺骗是把他人纯粹用作手段的两种方式。

但是假定在自卫情形的某种变体中，我阻止布朗杀我的办法是，以开枪来威胁他，或者告诉他“警察马上就会到”的假消息。尽管我会强制或欺骗布朗，但我可能没有把布朗看作纯粹的手段。我强制或欺骗布朗是因为这是唯一的方式，我既能阻止他杀我，又不伤害他。接下来假定，在

225

绝望困境情形中，我和你被困在海底的一个潜水钟中。尽管我们无法指望在十个小时之内获救，但我们的氧气只够两个人用6-7个小时。正如我所知，除非有一个人马上就死掉，否则两个人都会死。我开始以某种方式自杀，从而救你的命。当你试图阻止我时，我强制你或欺骗你，这样你就无法阻止我。

尽管我是在欺骗或强制你，但我没有把你纯粹当作手段。正如以前一样，我们若为了别人的缘故而牺牲自己的生命，则没有把他当作纯粹的手段来对待。

当奥尼尔解释“强制和欺骗是把他人纯粹当作手段对待”这个主张时，她写道：

把某物纯粹当作手段来对待就是以适合于对待物件的方式来对待。

然而，欺骗和强制并不是对待物件恰当的方式，因为这两者甚至都是不可能的。

科丝嘉也写道，基于康德的观点：

除了诉诸理性，企图以任何其他方式来控制他人的行动和反应都是把她当作纯粹的手段。

这个主张蕴含着，只要处于权威位置的人们告诉我们要做某事——诸如出示车票、填报税单或系安全带——他们就都是在把我们纯粹当作手段。这并不正确。科丝嘉还写道，基于康德的观点，只要“我们做某事能成功是因为大部分其他人不会做它”，我们就是把其他人当作纯粹手段。但是当穷人吃别人丢弃的残余物时，他们并没有把其他人当作纯粹的手段。

226

接下来我们假定，在

不助人为乐的情形中，当驾车经过沙漠时，我看见你受伤躺在路上，需要救援。我无视你，继续往前开。

按照某些作者的看法，康德会主张我在此把你当作纯粹的手段来对待。这个主张是错误的。我在无视你时，我没有以任何方式利用你，这样我就不可能是纯粹在利用你。

这些作者可能回应道，当康德用术语“纯粹当作手段”（或更准确地说，它的德语对应词）时，康德不是在其日常意义上使用的。康德常在特殊意义上使用这个语词。当我驾车经过你，忽视你的帮助需要时，真实情况可能是，我是在康德所想的某种特殊意义上把你当作纯粹的手段。奥尼尔和科丝嘉可能同样主张，所有的欺骗和强制确实在康德的特殊意义上把人们当作纯粹的手段。

我们会在非日常意义上使用某些语词，有时候这是有证成的。例如，我们值得扩大“痛苦的”意义，这样它能适用于诸如恶心这样不舒服的感觉。通过在这种宽泛的意义上使用“痛苦的”，我们就避免要不停地写上“痛苦的和不舒服的”，并且我们忽略这个区分一般不要紧。有时候不舒服的感觉远比某些痛苦更为糟糕。然而，在特殊意义上使用语词通常是错误的。我们于是可能会提出那种只是看起来重要而误导人的主张。例如，罗尔斯提出，如果我们接受他的契约论的道德理论，我们就应该使用“正当的”意指：符合由其虚拟的契约方所选择的原则。这使得“按照这些原则来行动是正当的”这个主张是无足轻重的。罗尔斯还提出，我们可以在这些原则会被这些契约方选择的意义上称它们为“真的”。这就使得“这些被选原则为真”这个主张是无足轻重的。

我们如果认为康德是在某种特殊的意义上使用“纯粹作为手段”，那就不应当说，按照康德的观点，我们必须从不把人们纯粹作为手段来对待。如果我们言尽于此，那么听众可能以为我们是在主张，按照康德的观点，我们必须从不把人们纯粹作为手段。为了避免误解，我们应该使用其他的术语。我们也许会说，基于康德的观点，我们必须从不以某种方式对待人们，我们可称此为把人们当作“纯单手段”（*shmerely as a means*）来对待。我们然后可来解释这种新术语的意指。

我认为，“纯粹作为手段”这个术语具有一种相当清楚且在道德上重要的日常意义。尽管康德可能有时候在特殊的意义使用这个术语，但我认为，他也

在日常意义上使用它。按照康德的人性公式，我们必须永不把人纯粹作为手段来对待，这样说是没什么误解的。这是康德最值得讨论的公式版本。

按照我对这种日常意义的粗略定义，我们若利用某个人，将其视作纯粹的工具，忽略其福祉和道德主张，对待他的方式就是要最好地实现自己的目标，则我们就是把他纯粹当作手段来对待。若要么（1）我们对这些人的对待在充分重要的意义上受到道德上相关的信念所管辖，要么（2）我们为了这个人的缘故确实或会在相关意义上选择承担某种大的负担，则我们就没有把他纯粹当作手段来对待，也没有接近如此。

当人们给出其他的定义时，他们常常是努力使得康德的主张涵盖大量的行动。我已经说过，我们有时候可以不使用特殊意义上的“纯粹当作手段”，而是修正康德的主张，以便它也谴责接近于把人们纯粹当作手段来对待。有时候我们应该做的不是拓展康德的主张来涵盖这样的情形，而是做出其他类似的主张。当不助人为乐者无视那些需要紧急救助的人时，他们没有把这个人纯粹当作手段来对待。但他们确实把这个人纯粹当作物件，没有任何重要性，就像路上的一块石头或是一堆破旧衣物。我们可以主张，这也是同样坏的。还有些对待人们的方式比把他们当作纯粹的手段更为糟糕。尽管希特勒在他征服的东部领域里把斯拉夫人纯粹当作手段，但他对待犹太人的方式则更糟糕。

228

32 作为手段的伤害

我们现在有这样一个问题：是否正如康德所主张的，不仅仅是把人们纯粹视作手段是不当的，而且以把人们纯粹作为手段来对待的方式来行动也是不当的。

正如我们已经看过的，康德的主张太强了。暴徒在买咖啡时把咖啡服务员纯粹作为手段来对待，但尽管这个人的态度是不当的，其行动却没有不当。当利己主义者冒着生命危险救溺水儿童时，尽管他纯粹是利用这个儿童作为得到某种报偿的手段，但其行动并无不当。

正如我说过的，为了回应这种反驳，我们能修正康德的主张。按照

第三种纯粹手段原则：如果我们以把人们纯粹当作手段的方式行动，或者接近于这样做，且他们还很可能受到伤害，那么这种行动是不当的。

但我认为，我们应当拒绝这个原则。让我们再次比较

救生艇情形，其中你能救我或者救另外五个人。

隧道情形，其中你能转变奔驰列车的轨道，它由此会杀死我而不是那五个人。

以及

桥梁情形，其中你只能杀死我来救那五个人。

按照某种观点，在这三种情形中你都应当救五个人。在救这五个人时，你是否会杀死我是不会产生差别的。当人们的生命受到威胁时，你应当做能救最多人命的任何事情。

229 按照第二种观点，只在救生艇情形中你应当救五个人的命。我们不杀人的义务优先于我们救人的义务。基于这种观点，在隧道情形和桥梁情形中，你救五个人都是不当的，既然这两种救五个人的方式会杀掉我。正如以前一样，你是否把杀人作为一种手段是会产生差别的。

按照第三种观点，在救生艇情形中，你应当救那五个人，而在隧道情形中，你至少是被允许去救那五个人的，但是在桥梁情形中，你救那五个人是不当的。我认为，这种观点是三种观点中最为人所接受的。基于这种观点，你是否把杀死我作为一种手段确实会产生差别。

如果我们接受第三种观点，我们可能会诉诸

有害手段原则：强加伤害给某人作为手段来实现某个目标是不当的，除非

(1) 没有更好的方式达到这个目标，且

(2) 给定这个目标的善性，我们施加的伤害不是不成比例的，或者不是太大的。

这个原则并没有告诉我们何种伤害会是太大的。这里我们不得不使用自己的判断。基于某种观点，我们能够有证成地给人们施加的伤害数量是有上限的。例如，按照汤普逊（Thomson）的讲法，不管我们能够救多少人的命，杀

死或严重伤害一个无辜者都是不当的。我们大多数人会接受一个不那么极端的观点。我们会认为，如果杀一个无辜者是能阻止原子弹爆炸的唯一方法，从而救下数百万人的生命，这就是正当的。但我们可能认为，把杀死某人作为手段只救五个人，或者仅仅是五十个人，这是不当的。在这之间会可能存着居间的情形，这个道德问题没有清楚或确定的答案。

按照我所称的标准观点，如果不经人们的同意，以伤害他们作为手段来实现某个目标，我们由此就把这个人纯粹作为手段来对待。正如我所论证的，这可能并不为真。我以伤害布朗作为手段来保护自己，但没有把布朗作为手段来对待，这样我就不可能把他纯粹作为手段来对待。

我们回到这样的情形，其中如果我们以施加伤害于某人作为手段，我们就可能也是把这个人作为手段来对待。当我们询问此种行动是否不当时，我们有两个问题：

Q1: 这个行动的不当性可能部分地取决于我们是否以伤害此人作为手段来实现某个目标吗？

Q2: 这个行动的不当性可能部分地取决于我们是否也把这个人纯粹作为手段来对待吗？

当我们比较诸如桥梁情形和隧道情形时，我们可能会判断对Q1的答案是肯定的。我们可能认为，尽管你能有证成地给奔驰着的列车转向，由此它会杀死我而不是那五个人，但你杀死我来救那五个人则是不当的。我一直都没有用论证反对这种观点。

我认为，对Q2的答案一直或近乎一直是否定的。如果你在桥梁情形中未经我的同意而杀死我，你可能没有把我纯粹作为手段来对待，或接近于如此。你对我的对待可能在充分重要的意义上受到相关道德原则（诸如康德的同意原则）的管辖。情况可能是，如果你本来就靠近这个电车，你会牺牲自己而不是我来救那五个人。但我认为，这些事实不会影响你是否行动不当。如果你牺牲我作为救那五个人的手段是不当的，那么不管你是否还把我纯粹作为手段来对待，这都是不当的。即使你没有把我纯粹作为手段来对待，或接近于这样做，这些事实也不能证成你的行动。

我们接下来转向这样的情形，其中我们以伤害某人作为手段是有证成的。在第三种地震情形中，你只能不经布莱克的同意粉碎其脚趾头来救你的孩子。

我认为，这个行动是能得到证成的。如果某人为了救其孩子而粉碎了我一个脚趾头，（我希望）我不会抱怨。有些人尽管会认为该行动不当，但还是会接受：我们可以有证成地将某些较小的伤害强加给某人，如果这是救另一人生命的唯一方法。例如，按照汤普逊的看法，为了救某人的命，我们可以擦伤别人的腿，使得这个人具有“短期的轻度疼痛”。这样我们能够假定，在

第四种地震情形中，暴徒只能在不经布莱克的同意情况下擦伤她的腿，使她具有短期的疼痛，由此来救自己孩子的命。

这个暴徒视布莱克为纯粹的手段。如果有助于实现他的任何目标，他会杀死或重伤布莱克。这样，如果暴徒以擦伤布莱克的腿来救其孩子的命，他就既施加伤害于布莱克，也把他作为纯粹的手段来对待。按照康德的人性公式（这包括纯粹手段原则），以任何把人们纯粹当作手段来对待的方式都是不当的。按照第三种纯粹手段原则，以任何把人们纯粹当作手段来对待且伤害他们的行动方式都是不当的。这些原则都蕴含着，暴徒如果擦伤布莱克的腿救其孩子，将是行动不当。

232 这个结论是让人无法接受的，尽管这个暴徒对布莱克有不当的态度，但可以有证成地给布莱克造成较小的伤害来救其孩子。这个孩子具有得救的道德要求权。这种要求权不会受到其父亲对布莱克态度的不当性所破坏或否决。这也同样适用于其他情形。如果在第三种地震情形中，在道德上允许你使布莱克失去一个脚趾头来救你的孩子，那么在道德上也允许这个暴徒以同样的方式救他的孩子。

人们广泛认为，为了解释以伤害某人作为利他手段的不当性，我们能诉诸康德的这个主张：我们必须从不把人们纯粹作为手段来对待。我已经论证，这个信念是错误的。如果以伤害某人为手段来实现某个好目标是不当的，那么即使我们没有把这些人纯粹作为手段来对待，这些行动依然是不当的。如果以施加较小的伤害某人为手段来实现某个好目标不是不当的，那么即使我们是把这些人纯粹作为手段来对待，这些行动依然不是不当的。

康德的主张包含着重大的真理。我们把任何人视作纯粹手段都是不当的。但我们行动的不当性从不或很少取决于我们是否把人们纯粹作为手段来对待。

第十章 尊重与价值

33 对人 (persons) 的尊重

康德在对人性公式的另一个评论中写道：

233

每个理性存在者……必须总是被视作一个目的……是尊重的一个对象。

尊重一切人的要求是康德对我们道德思考所做的最伟大的贡献之一。但它并没告诉我们应该如何行动。

伍德 (Wood) 表示，

(A) 我们应该总是以表达尊重的方式对待人们。

然而，我们无需对人们表达尊重，就能正当地对待他们。伍德表示，只要我们正当地对待人们，我们的行动就可看作对这些人表达了尊重。但是基于这个建议，(A) 告诉我们的只是我们必须总是正当地对待人们。(A) 不会有助于我们确定哪些行动是正当的，因为我们无法判定某个行动是否对人们表达了尊重，除非通过判定某个行动正当与否。

有些作者表示

(B) 如果对待人们的方式与对他们的尊重是不相容的，这就是不当的。

某些不当的行动显然与对人们的尊重是不相容的。康德的例子是：耻辱性的或羞辱性的惩罚、嘲弄、诽谤，以及展示傲慢或鄙视的行动。但是康德的公式意图涵盖一切不当的行动，而大多数不当的行动并不以这种不尊重的方式对待人们。

有些作者表示，一切不当行动都在广义上与对人们的尊重是不相容的。按照这个建议，(B)不会是一个有用的主张。如同以前，为了决定某个行动是否在这种广义上与对人们的尊重不相容，我们首先就要判定这个行动是否是不当的。如果这个行动不是不当的，它就会与对人们的尊重是相容的。正如康德与西季威克所提醒的，道德哲学家通常会提出那种看似会给我们“有价值信息”的主张，但实际上只告诉我们：若某些行动是不当的，则它们就是不当的。

康德还主张：

(C) 我们必须总是尊重人性，或者尊重使得我们成为人 (persons) 的那种“理性本质”。

伍德称(C)是康德的最高道德原则的“最有用的表述”。伍德主张，尽管(C)不能解决一切道德问题，但这个原则提供了“判定道德问题的正确基础”。为了支持这个主张，伍德指出，在其最后一本也是篇幅最长的一本有关道德的著作中，康德常常做出似乎是在诉诸(C)的论述。

我认为，康德的论述并未表明(C)是一个有用的原则。正如伍德本人所承认的，康德对(C)的诉诸通常是“简短且不经意的”。这个评论对康德的观点所增无几。例如，康德写道，我们发展自己天赋之责，“与内在于我们本人的人性目的息息相关”。康德还提出了伍德正确地拒绝的其他主张。康德主张，任何人给予自己性快感或者为了避免痛苦而早死，这都是不当的，因为这贬低或玷污了人性。康德在谴责撒谎——即便是“为了实现真正的好目的”的撒谎——的时候写道，任何撒谎者“违反了自己的人格尊严”，这样他变成了“纯粹虚有其表的人”，其具有的“价值甚至不如一个纯粹的物件”。并不是这些主张使得康德成了自古希腊以来最伟大的道德哲学家。

伍德表示，康德在提出这些主张时，他误用了(C)。伍德写道，我们能够拒绝康德有关性、自杀和谎言的主张，因为“我们能有证成地认为，在这些事情上我们对于尊重人性要求什么具有更好的了解。”在深刻的道德分歧的两

面都能使用这个原则来阐明他们认为是其最强大的论证，正是这个原则的“一种优势”。

这个评价在我看来是错误的。当康德主张某个行动会违背或贬低人性，以及我们拒绝这些主张时，无论是康德还是我们都没有给出我们最强大的论证。

(C)也无法帮助我们判定在困难的情形中，何种行动会是不当的。

34 两类价值

当康德解释“我们必须总是把理性存在者作为目的来对待”的意思时，他主张这类存在者具有尊严，这里尊严意指一类最高的价值。这个主张提出了伦理学上最深刻的问题之一：善如何与正当（或与我们在道德上应该做什么）相关联。康德还主张，我们不是要遵循古希腊人，即首先询问什么样的目的是善的，然后得出什么是正当的结论，而是要把这个程序颠倒过来。罗尔斯称康德道德理论的核心特征是“正当以这种方式‘优先于善’”。但相对而言，伍德主张，尽管康德的人性公式“采取了规则或命令的形式，但在基本上它断言的是一种实质价值的存在”。赫尔曼表示，康德的“最根本的理论概念”是“善”，康德的伦理学最好是理解为价值的伦理学。

在我们考虑康德的价值主张之前，做出更多的区分是有益的。许多事物是在我所称的蕴含理由的意义上是好的或坏的。这样的事物具有某类属性或特征，它们在某些情境下给予我们或他人理由对这些事物以某种方式做出回应。

某些这样的好事物具有一类正如斯坎伦和其他人所说的有待促进的价值。两个这样的例子是幸福以及痛苦的减轻或防止。当事物具有这类价值时，实际上正是这些事物而不是其价值，才是我们有理由去促进的。

236

我们能够促进的是事件，广义的“事件”也能涵盖行动和事态。事件能够作为目的或作为实现目的的手段而是好的或坏的。基于某些观点，行动只能作为手段而是好的或坏的。我认为，我们应该拒绝这样的观点。例如，我们若把孩子抚养好，作为好的爱人或友人表现良好，成功地从事各种各样有价值的活动，或者正当地行动和尊重地对待他人等，我们就行为得体。这样的事情——不仅作为实现幸福或其他好目的的手段，而且部分或完全为了自身的缘故——都可能是值得做的。这样我们就应该把行动纳入那些作为目的就可以有好坏之分的事件之中。

按照在我看来对事件善性最好的观点，我称之为

实现主义 (actualism)：某些有可能的行动或其他事件如果具有一些内在属性或特征——这些属性或特征给予我们理由，从而想要通过采取该行动或使该事发生而让它们成为现实且尽可能使之变成现实——那么它们就作为目的是好的。有可能的行动或其他事件如果变成现实是实现某个目的的有效方法，那么它们就作为手段是好的。

这同样适用于作为目的而坏的事件，或作为实现某个目的之手段而坏的事件。事件可能是对特定的人或在蕴含不偏不倚的理由的意义上（或两者）作为目的而是坏的。我们除了有理由设法导致好的事件和防止坏的事件，我们还有理由对它们持有各种各样的态度，诸如希望、高兴、恐惧和遗憾。所有这些态度都是针对“这样的事件是现实或真实的”的可能性或事实，并作为事情进展状态的一个组成部分。

既然“实现主义”适用于一切可能的行动和它们所有可能的效果，这种观点涵盖所有这样的事物，其善性直接与我们应该做什么的任何决策直接相关。我们有理由以某种方式行动，当且仅当或正好当这个行动以这种方式要么作为目的是好的，要么是作为实现某个好目的的手段而是好的。然而，实现主义并不主张要涵盖不是行动或其他事件之事物的善性。

按照某些作者的看法，这种观点可以拓宽，从而涵盖某些持续存在的事物（诸如人或艺术品）的善性。某些事物的性质若给予我们理由想要它们存在或持续存在，并且我们有理由尽力那样去做，则人们主张这些事物是好的。摩尔甚至写道：

当我们断言某物是好的时候，我们的意思是它的存在或实体就是好的。

但这些主张是错误的。尽管某事物本身不是好的，但其存在可能是好的，反之亦然。例如，有许多坏人，但他们的存在作为目的是好的。当某些好人正要极痛苦地慢慢死掉的时候，他的持续存在作为目的就是坏的。如果好的艺术品从未有人见过，那么它们的持续存在就没有任何价值。

按照斯坎伦所称的目的论理论，只有行动或其他事件才具有在自身就是好的意义上的内在价值。斯坎伦正确地拒绝了这个主张。还存在其他可能自身就好的事物，诸如人、书和论证。既然这些事物不是事件，我们就无法想要它

们发生或使得它们发生。但我们可以以其他方式回应它们。我们有理由读这些书，被好的论证所说服，尽力变得更像好人。

我们现在转向另一类价值，正如斯坎伦和其他人所说，它是要被尊重而不是被促进的。如以前一样，当事物具有这种价值时，确实是这些事物而不是其价值是我们有理由去尊重的。我们尽管可以主张人是具有这种价值的最好例子，但可以从其他的例子开始。这些事物可以是那种被宣称为具有象征、历史或联想价值的东西，诸如国旗、活古树、肖像和其他宗教画以及死者躯体等。

斯坎伦写道，理解某物的价值在部分程度上就是“知道如何评价它（知道要求什么样的行动与态度）的问题”。许多这样的行动和态度可以不严格地被称为尊重或尊敬该事物的方式。我可能尊重国旗、古树和宗教画，从而拒绝用国旗作抹布，用最老的树作柴火和用宗教画作标靶。为了对许多这样事物的价值对出恰当的回应当，我们应该保护它们，让其继续存在。但情况并非总是如此。我们对死者躯体的价值要做出恰当的回应，不是要如古埃及人那样保存它们，而是以某种恭敬的方式毁灭它们，如放在装点着鲜花的火葬柴火中烧掉，而不是把它们扔在垃圾堆之中。

这类事物的价值完全不同于好目的或好人的善性。它不是一种善性。尽管有些亡者的躯体可以作为尸体，供解剖学或手术教学之用，其他的躯体可以在某些恐怖电影中用作道具而是好的。这些并不是所有的死者躯体都可以被宣称具有的那类价值。有些宗教画并不是好的。尽管这类价值不是一种善性，也不是有待促进的价值，但当我们以尊重的方式对待它们来回应这类事物的价值时，这些行动可以作为目的而是好的，具有那种有待促进的价值。

我们接下来可以转到有关人的生命价值的主张。斯坎伦写道，欣赏这种价值：

首先是把人的生命视为某种应予尊重的东西，其中包括要看到不摧毁它们、保护它们以及想要它们进展顺利的各种理由。

然而，我们不必为了看出我们拥有这些理由，而把人的生命视为具有应予尊重而非有待促进的价值。当人们活得好的时候，在蕴含理由的意义上这既对这些人好的，也是客观地好的。这种幸福的生活是作为目的而好的。我们有理由保护美好的生活且让其生活在其他方面进展顺利。

按照某些观点，人的生命还有另一种价值。假定你要开始一种痛苦而没尊严的缓慢死亡，你又没有任何重要的事情可做。你可能就有极强的理由自杀，

其他人有很强的理由帮你这样做。在那些诉诸人的生命价值的人之中，有些人会认为这样的行动会是不当的。这些人可能会同意，你早点自然地死亡对你好的，也在非个人的意义上是好的。但这些人认为，你不应该自杀，其他人也不应该帮助你，因为这样的行动没有尊重人的生命价值。基于这种观点，尊重个人生命的价值不同于，甚至冲突于去做既对这个人是最好的事情也是这个人会选择的事情。

斯坎伦拒绝这种观点。他写道，我们有理由不去结束某个人的生命，仅仅是“只要这个人的生命是他有理由继续活下去或想要去活的”。斯坎伦这里否认了人的生命具有这样一类价值，我们对它的尊重方式与这个人的福祉与自主性相冲突。我认为，这是有关人类生命价值的正确观点。为了捍卫在这样的情形下自杀和帮助自杀是不当的，我们还需要其他的论证。

不是人的生命，而是活着的人才应该被宣称拥有那种我们应该尊重、而非促进的价值。斯坎伦主张，我们应该尊重这种价值，只以能向这些人证成的方式对待他们。康德做出了类似的主张，要尊重人们，我们应该只以他们能够理性地同意的方式对待他们。

35 康德式尊严

我们接下来考虑康德有关价值的主张。康德在提出这些主张时，区分了三类目的。康德所称的有待产生的目的是我们能努力实现或产生的目标或结果。这些是日常意义上的目的，正如“减轻痛苦是好目的”这种主张中的那种意义。康德把这种目的与现存的或既存的目的做了对比，后者的例子主要是理性存在者或人。康德的第三种目的是自在目的。这类事物具有康德所称的尊严，他将之定义为绝对的、无条件的和无可比拟的价值（worth or value）。在别无他物具有更高价值的意义上，这种价值是最高的或者无可超越的。

240

按照某些作者来说，康德认为这种最高价值只为某些现存目的，诸如理性存在者所具有，其价值是有待尊重而不是要被促进的那种。但是康德主张，存在几种有待产生的目的具有最高价值，这些目的是我们应该努力促进或实现的。

这样的一种目的是拥有善良意志。康德主张，若我们之所以履行义务，是因为它们是我们的义务，而不是由于某种其他的目标如避免惩罚，则我们的意志就是善良的。我们对善良意志的拥有要么会被看作是一种精神状态或倾向，要么是构成善良意愿的一种活动。基于康德的观点，无论是以哪种方式，拥有

善良意志都是我们应当努力达成之事。以康德自己的话来说，“理性的真正天职必定产生善良的意志。”

另一种有待产生且具有最高善性的目的是康德所称的目的王国。这是一种可能的事态或可能的世界，若每个人都具有善良意志且总是行动正当，我们就会一起产生那种世界。

第三种这样的目的是康德所称的最高或最伟大的善。这种可能世界是目的王国再加一个特征：每个人都拥有与其德性相配的一切幸福。康德主张“我们应尽力促进”这个目的，“理性……命令我们竭尽所能来产生它”。

可能还有第四种这样的目的。康德称理性存在者是“其存在本身就具有绝对价值”。他写道，如果不存在任何理性存在者，宇宙将是“纯粹的浪费、徒劳，没有终极的目的”。这些评论显示，基于康德的观点，理性存在者的持续存在是带有最高价值的、有待产生的另一种目的。

我们可以回到康德这个主张：理性存在者或人是自在目的，具有尊严或最高价值。正如我说过的，人不是有待产生的目的。他们的价值是不同类的。正如伍德和赫尔曼主张的，基于康德的观点，“即使最坏的人也有尊严”，一个具有善良意志的人与一个具有普通意志或坏意志的人具有“的价值一样多”。我认为，康德观点的这个部分是深刻的真理。但是道德上最坏的人所具有的价值不是一种善性。希特勒和斯大林不是好人。人们具有尊严或价值，这是在如下极为不同的意义上讲的：鉴于其作为理性存在者的本质，必须总是以某种有益或尊重的方式来对待他们。我认为同样的主张适用于一切有感知的存在。在一种经拓展的康德式意义上，即使是能够感知痛苦的最低级的蠕虫也有尊严。一只蠕虫不可能自身是善的，但是其本质使得它是这样一种存在，我们若给它施加无意义的痛苦就是不当的。

241

我一直在忽略一种复杂性。康德有时候使用“人性”来指称理性，他也称为“理性本质”的东西。这样，当康德主张人性是一种具有尊严或最高价值的自在目的时，他可能意指理性具有这种价值。尽管理性存在者的价值不是一种善性，但可以主张，他们是理性的从而是善的。赫尔曼写道，在康德的伦理学中，“‘善的’领域是理性的活动与能动性”，康德把道德建立在“作为一种价值的理性”的基础上。伍德甚至称康德有关理性价值的主张是“康德整个伦理学理论中最为根本的命题”。

基于康德的观点，和拥有善良意志一样，理性在部分程度上是有待产生或促进的一个目的。我们应该运用我们的理性，并且通过发展我们的理性能力而

努力变得更理性。康德称尊严是一种价值，无限地高于一种低等的价值，后者他称为价格。在只具有价格的事物中，康德纳入了快乐与没有痛苦。这样，如果康德想主张理性或理性的活动具有尊严，那么他的观点将蕴含理性具有的价值无限地大于消除痛苦具有的价值。纽曼（Cardinal Newman）主张，尽管罪与痛都是坏的，但是罪是无限地坏的，乃至如果全人类都处于极度痛苦之中，这还没有犯下一种可宽恕的罪那么坏。尽管这种观点很可怕，但我们能够理解为什么有人会持有它，因为我们能够明白罪看起来如何可能无限地比痛苦糟糕。如果说理性和理性活动具有无限价值意义上的尊严，防止痛苦只有有限的价值，那么康德的观点就具有甚至更难以让人接受的含义。例如，按照这种观点，我们应当增强下棋的能力，或者解决十字谜的能力，而不是使无数人免于极度的痛苦。这个结论是疯狂的。

242 人们可能反驳道，即使基于这种观点，我们也应该免除这些人的痛苦，因为这会有助于他们的行动更为合理。但是我们可以在外科手术期间通过使这些人失去知觉的方式免除他们的痛苦，而这将无助于其行动的合理性。

有人接下来可能会主张，理性的价值是有待被尊重而不是被促进的。这并非康德的观点，因为康德通常主张我们应当努力发展和使用我们的理性能力。而且这种修正版本的康德观点还是会面临类似的反驳。我们尊重人的价值，不是通过给这个世界增加人口，而是通过遵循各种各样其他的道德要求，诸如不杀人或害人。如果正如希尔指出的，理性具有这种价值，那就应该存在类似的要求不去破坏或影响人们的理性能力。如果理性的价值无限地高于一切价格，那么为了只有价格的事物（例如消除痛苦）而“牺牲”或“交易”任何理性能力都是不当的。于是，不管为多少人免除极度的痛苦，只要这会损害我们下棋的能力或解决十字谜的能力，就是不当的。这也是个非常愚蠢的结论。

我认为，康德的观点并不具有这种含义。当康德主张人性具有尊严时，我认为他很少指涉理性。康德区分了下述两者：（1）我们道德地行动和具有善良意志的能力和（2）我们其他的理性能力。我们可以称（2）为我们的非道德的理性。在刚刚把尊严定义为一类绝对和无可比拟的价值之后，康德就写道

道德和人性（就其作为道德能力而言）是独立地具有尊严的事物。

“人性”这个词在此不可能指称非道德的理性。康德在其他许多段落中区分了我们自己和他所称为的“我们中的人性”。我认为，“人性”的这种用法绝

大部分不是指称我们的理性，而是要么指称我们道德地行动和具有善良意志的能力，要么指称作为康德所称为的自体（noumenal）存在的我们自己。尽管康德的有些评论显示非道德的理性也是具有最高价值的自在目的，但我认为，他并没有承诺这种观点。康德是“伟大思想家中最不精确的一位”，他对“人性”这个词的用法不断变换且含糊不清。康德确实谴责某些恶习，如贪吃和醉酒，其根据就在于这种恶习干扰了我们理性能力的使用。但康德的主要主张并不蕴含着：吃得太多和醉酒是不当的——这即便是使无论多少人从无论多剧烈的痛苦中摆脱出来的唯一方式，也是不当的。

赫尔曼写道，康德有关价值的主张提出了“对传统观念的一种彻底批判”。基于康德的观点，“过去的道德哲学……搞错了善的本质”。

我认为，康德并没有提出这样的一种批判。如果康德主张没有任何事物具有那类有待促进的价值，他就会拒绝早前的诸多观点。但是我们已经看到，康德主张，目的王国、拥有善良意志、康德的最伟大的善（每个人都有德有福）的那种可能事态或可能世界都拥有这种价值。基于康德的观点，这些都是有待产生的目的，是我们应该尽可能地促进的。康德在主张何种事物具有这类价值时，他也追随了此前的哲学家，他们许多人都主张，德性和幸福就是两种作为目的而善的事物。

康德可能不接受一种广泛持有的价值观，因为康德通常忽略有些事物能在蕴含理由的意义上是非道德地或好或坏的。例如，他主张，审慎原则，或者做促进我们自己幸福的事情的原则仅仅是假言命令，只是因为我们想要幸福，它才适用于我们。康德在此忽略了我们想要幸福的非道德理由。在其实践合理性的论说中，康德描述了道德和工具合理性，而两者之间几乎就是荒漠。然而，康德对非道德的善性的忽略（我会在附录G中讨论）并不是一种批判。

康德可能不会接受另一种被广泛持有的价值观点。基于此种观点，要是有价值的总是要在某种意义上是好的。当康德主张所有理性存在者都有他称为尊严的那类价值时，他并不意指所有理性存在者都是好的。如我所言，康德的意思是一切理性存在者都具有某种应予尊重的价值，因为应当仅以某些方式来对待他们）。这种价值是一种身份（status），或者赫尔曼所称的“道德地位（moral standing）”。这种价值被许多传统观点所忽略。

我认为，康德的这一主张是正确的：即便是道德极恶者也具有与其他任何人一样的道德身份。康德称这种身份为尊严或最高价值，从而以一种颇具说服力的方式表达了这个主张。要不是“道德身份”这种观念所具有的理论益处，

那就必须通过从更大群体的成员中挑出那些满足进一步条件的人，来做出某些区分。做个类比来说明这一点。在罗马法中，只有那些不是奴隶的人才有充分的法律身份，算作是人。在民主国家中，只有成年人人才具备有资格投票的身份，在许多国家只有那些是公民的人才有资格得到某些利益。相对比而言，基于康德的观点，所有的理性存在者或人都应该只受某种方式的对待。所有的理性存在者或人都具有这样的道德身份，即作为应当仅以某些方式受到对待的实体——这样的说法就几乎没增加任何东西。

康德的价值主张还部分地是误导的。正如我说过的，当康德主张所有的理性存在者具有尊严或最高价值时，他并不意指所有此类存在都是好的。但是康德主张，道德、善良意志、目的王国这种可能世界以及最伟大的善都拥有这种最高价值。基于康德的观点，这些事物的价值是一类善性。这样，康德的价值主张并没有区分最高的善与道德身份，而道德身份是极恶的存在者（如希特勒和斯大林）也具有。然而，我们可以很容易为康德补上这个区分。

36 正当与善

康德主张，最高的或最伟大的善是这样的一个世界：其中每个人既是完全有德或说道德上善的，且享有与此德性相配的一切幸福。康德还写道：

每个人都应该努力促进这种最伟大的善。

道德法则命令我们把世上最伟大的、有可能的善作为我们一切行动的终极目标。

245 按照我们可称之为的

最伟大善的公式：每个人都应当总是努力促进一个具有普遍德性及与之相配的幸福的世界。

这个理想的世界可能极难实现。因此，在应用这个公式时，我们应该与不完美却更容易实现的世界状态相比较，并问问怎样才能尽可能地接近康德的理想。

康德主张，如果每个人的德性与其幸福相称，或说配得上其幸福，这是最好的。在理想世界里，情况确实会是：我们所有人都是有全德全福的。有些作

者表示，在那些不理想的世界里，最好的世界是满足两者相称的世界。但这看来不大像康德的观点。如果没人有德性和幸福，或者每个人都有罪且悲惨，其德福也是相称的。但这样的世界显然比下述世界要差得多：其中每个人都很有德性且很幸福，只是有些人的幸福与其应得稍有出入。这样我们可以假定，基于康德的观点，若人们的德性和幸福更多，那么即使两者不那么相称，这样的世界始终也是更好的。

康德的如下主张是不可行的：任何人都不能影响其他人的德性。基于这个假设，我们只能通过提高自己的德性水平来促进德性。我们只能通过具有善良意志和做任何我们该做的事情，由此最好地促进德性。我们只能把幸福给予那些幸福没有应得的那么多的人，从而最好地促进应得的幸福。常常有人主张，我们无法这样做，因为我们不知道人们应得多少幸福。然而，我们并不需要这种知识。我们只要对于“哪些人更有可能应得更多幸福”有理性的信念就足够了。正如康德所假定的，我们常常具有这种信念。通过设法使这些人更加幸福，我们可以按照这种信念来行动。这样，康德的最伟大善的公式给了我们一个可以努力去实现的目标。

我们接下来可以再做些区分，引入康德的其他一些主张。道德理论在最广泛的意义上是

行动后果论的，如果它们主张每个人应该总是去做或努力去做任何能最好地实现一个或更多共同目标的事情。

按照一种这样的理论，享乐主义的行动效用主义是

HAU：每个人应该总是产生或努力产生最多（减掉痛苦之后）的幸福。

这些理论是个人中立的（**person-neutral**），意思是他们给予每个人完全一样的共同目标。按照大多数道德理论以及大多数人的道德信念来说，存在一些每个人都应该努力实现的目标，比如人们不应挨饿。但是我们每个人都还应该努力实现许多个人相对的（**person-relative**）道德目标。基于这样的观点，例如，我们不是具有如遵守诺言和照看孩子这样的共同目标，而是我们每个人应该信守自己的诺言和照看自己的孩子。第三组观点则并不给予我们任何

共同的道德目标。例如，“我们唯一的义务就是服从十诫”这样的观点就是如此。

有些道德理论是全部或部分地基于价值的，意思是它们诉诸这样的主张：什么东西在重要且实质的意义上是好的或坏的。按照我们所称的基于价值的行动后果论，或者

VAC：每个人应该总是去做或努力去使得事物进展最好的任何事情。

例如，基于这种版本的HAU，每个人应该产生或努力产生最大净数量的幸福，因为这是我们如何使得事物进展最好的方式。

247 有些道德理论除了提出“什么是善的”和“什么是我们道德上应当做的”主张之外，还会提出善这个概念与道德版本的“应当”概念如何相关联的主张。按照某些理论，“善”这个概念是根本的，可以用来定义这种版本的“应当”概念。按照其他的某些理论，“应当”概念才是根本的，可以用它来定义善这个概念。按照第三组理论，这两个概念无法依据彼此来定义。我认为，第三组理论是最好的。因为只有这些理论才在独立的意义使用“善”和“应当”，只有这些理论才能对“什么是善”和“道德上应当做什么”之间的关系提出实质性的主张。

我们可以把摩尔的《伦理学原理》作为第一类理论的一个例子。摩尔主张，当我们说

我们应当做某事，或这个行动是正当的，我们意指这个行动使得事物进展最好，由此产生最多的善。

我们可以称此为促进善性意义上的“应当”。摩尔还主张：

M1：每个人的行动应该总是使得事物进展最好。

这个主张可能看来就是一种基于价值版本的行动后果论。但是如果摩尔是在其促进善性的意义上使用“应当”，则M1是个隐藏的同义反复，其公开的形式之一是

M2: 如果每个人总是做使得事物进展最好的行动，则每个人总是会做使得事物进展最好的行动。

每个人都应该接受这个主张，不管其道德信念如何。摩尔的《原理》并没有提出一个实质性的道德观点。

康德主张我们应该根据“应当”来定义“善”，所以康德的观点与摩尔的恰恰相反。用康德的话来说

善恶的概念绝不可先于道德法则而被决定……而是只能在其后……并且借助于它。

令人吃惊的是，康德还主张：

所有的命令都是以“应当”表达的……且说……某个行动会是善的。

康德这里所做的似乎恰恰是他主张我们绝不能做的事情，即根据善来定义应当。康德同样称某种行动“实践上必然的，也就是好（善）的”。但这些评论并没有在其任何日常意义上使用“好的”。例如，在这些日常意义上，某些行动是好的，尽管其他的行动可能是更好的。在这些和其他的段落中，康德并没有区分“某些行动是好的”与“这个行动是实践上必然的”，即我们应当做的。后面那些语言才能更好地表达康德心中所想。我由此提出，当康德称某个行动是“好的”时候，他意指这个行动是我们应当做的。既然康德是在“基于应当”的意义上使用“好的”，于是他遵循了自己的要求：“善（好）”要根据“应当”来定义。

248

当康德称某个目的或结果是“好的”或“最好的”时候，他看来通常使用的是一种基于应当的意义。例如，当康德主张

K1: 善良意志是最高的善。

他似乎在部分程度上意指：

K2: 每个人都应当努力拥有善良意志。

但康德也可能意指：我们之所以应当努力拥有这样的意志，是因为这种意志是至上的善。“善”的这种用法不会是基于应当的。在这个方面，康德的道德理论可能正如赫尔曼所主张的是种价值伦理学。但是，在从其关于何者为善的信念之中导出道德法则的内容之时，康德不会是在做那种他宣称我们绝不可做的事情。从“善良意志是至上的善”这个主张，我们能推出K2。但是对于我们应当做什么，我们无法得出任何其他的结论。

康德宣称，古希腊人确实犯了这种错误，因为他们试图从有关至善或最伟大的善的信念中推出道德法则。然而，如我们之所见，康德本人描述了一个他称为最高的或最伟大的善的理想世界，他主张每个人应该努力去产生这个世界。康德在此犯了他所说的古希腊人的“根本错误”吗？他是从有关最伟大的善的信念中推出“我们应当做什么”的信念吗？

情况或许如此。如我们之所见，康德主张：

K3：每个人应当总是努力促进最伟大的善。

这看来似乎是另一版本的基于价值的行动后果论。康德可能看来是在主张，每个人应该总是努力产生最好的或是最伟大善的世界。他还说了其他这样的话语，当他写道：对于每个人而言，“他在每个场合的义务就是在其能力范围之内做一切善的事情。”

我认为，这不是阐释K3的最好方式。我认为，康德是在基于应当的意义上使用了“最伟大的善”这个术语，意指“每个人应当始终努力促进的事情”。如果这就是康德的意指，K3就可以重述为：

K4：每个人应该总是尽力去促进那种每个人应该总是尽力去促进的世界。

这个主张看起来是一个纯粹的同义反复，这是每个人都接受的。但情况并非如此。K4蕴含着我们应当接受某种版本的行动后果论，因为K4蕴含着存在着某种世界，它是每个人应当始终尽力去促进的。许多人会拒绝这个主张。

然而，K4并不蕴含一种基于价值版本的行动后果论。当康德主张K3时，他可能也是用“最伟大的善”指称某种可能的世界，他在别处宣称这种世界是最伟大的善。K3由此可以更充分地陈述为：

K5: 每个人应该总是尽力促进一个普遍具有德性和应得幸福的世界。

这是对于康德这部分观点的最为清楚的陈述，这个主张甚至并没有使用“好的”或“最好的”这种语词。这样，康德的行动后果论版本在很大程度上并不是基于价值的。

37 促进善

康德的观点也并不显然是行动后果论的。康德的最伟大善的公式可能被宣称是我们所需要的唯一原则，因为我们应该总是尽力直接地促进康德的理想世界。但这并不是康德的观点。康德主张我们应该遵循某种其他的公式，诸如他的人性公式和普遍法则公式。这样我们接下来询问：康德有关最伟大善的主张如何与他的其他公式相关联。

康德写道，

我们能够假定对道德法则的履行会导向最高的目的。

他还写道：

对道德法则最严格的遵守，应被视为带来（作为目的的）至善的原因。

在这些和其他的段落中，康德假定

K6: 正是通过遵循（康德的其他公式所描述的）道德法则，每个人能够最好地促进最伟大的善。

每个人如果都遵循道德法则，具有善良意志，那么会因此促进康德理想世界中的一个要素，即普遍德性，因为这种普遍德性就在于每个人遵循道德法则和具有善良意志。但这并不是康德的全部意思。康德主张，如果每个人都遵循道德法则，这会导致最伟大的善或成为其原因，此时康德必定是指这个理想世

界中的另一个要素：普遍应得的幸福。因此康德似乎是假定：

K7：正是通过遵循道德法则，每个人能够最好地给予每个人以其德性相配的幸福。

251 尽管每个人遵循道德法则会使得这个世界更接近康德的理想，但康德主张，这并不足以充分实现这个目标，因为我们无法给予每个人以其应得的一切幸福。例如，有的好人年轻就死掉了。但我们能指望我们的灵魂不朽，并且我们死后，上帝会给予每个人以余下的应得幸福。

我们可能会怀疑康德是否能假定K7。康德似乎认为，我们应该遵循某种严格的规则，诸如禁止撒谎、偷盗和违背诺言。康德看起来不大可能相信遵循这样的规则能最有效地促进应得的幸福。

然而，那并非不大可能的。在康德的生活时代以及之前，人们广泛认为，

(A) 正是通过遵循常识规则，而不是试图直接地促进每个人的幸福，每个人才能够最好地促进每个人的幸福。

正如西季威克后来所论证的，这个假设也是相当可行的。在试图预测何种行动能产生最多的幸福时，人们会犯下严重的错误。他们通常以有利于己的方式自欺。例如，人们容易相信，对于能够偷到的财产，我们比其所有者更有需要。如果每个人总是尽力最大化幸福，这也会破坏或弱化各种有价值的社会实践和制度，诸如包含信任的诺言实践。如果每个人都具有那些总是试图最大化幸福的人所具有的那些动机，这在几个方面是坏的。为了能总是以这种方式行动，我们大多数人都不得不失去太多的动机，诸如对特定的人的强烈的爱，而我们的许多幸福正取决于这些动机。

我们接下来能够做出一些早期思想者并没有做出的区分。我现在用“后果论”只指称基于价值的观点，我将把“最好的或期望上最好的”简称为“最好的”。如果我们假定每个人会尽力遵循某组规则，那么某些可能的规则会是

最优的，其意思是如果这些是每个人尽力遵循的规则，事情就会进展最好。

西季威克出于刚才给出的理由，认为常识道德规则相当接近最优的规则。按照一种版本的规则后果论，或者说

RC: 每个人应该总是努力遵循最优的规则。

按照某种版本的行动后果论，或者

AC: 每个人的行动应该总是使得事情进展最好。

在那些接受这两者之一的人之中，大部分人现在假定这两者是相冲突的，这样我们必须二者择一。这些人认为

(B) 在某些情形中，违背某些最优规则很可能或肯定使得事情进展最好。

西季威克作为一个行动后果论者，他主张在这样的情形下，我们应该违背最优的规则。按照大部分规则后果论者的看法，我们应该转而遵循最优的规则，即使这样做，我们可能或肯定会使得事情进展得更糟。

然而，有些人拒绝(B)。这些人相信

(C) 正是努力遵循最优规则，每个人才总是最有可能使得事情进展得最好。

摩尔接近于接受(C)。摩尔主张，我们在努力做最有益的行动时应该总是遵循某种最优的常识规则。如果(C)为真，后果论的这两种形式就不会冲突而是相一致，我们能够同时接受两者。按照我们可称之为的行动-规则后果论，或者

ARC: 每个人应该总是努力遵循最优的规则，因为这是每个人最有可能使得事情进展最好的方式。

我们若询问(C)是否为真从而这两种形式的后果论相一致，就必须诉诸

应该如何评价我们行动效果的观点。按照我们称为的

253 边际主义观点：要判定一个行动有多好，我们应该问这个行动引起的差别是什么。该行动会产生的利益是：与不做它相比，如果做它，事情的进展就会更好的那部分余额。

当我们考虑某类情形时，这种观点看来是讲不通的。这类情形的一个例子是：只要一定数量的人做出某个行动，就会完全获得某种好结果。若这样做的人超过这个数量，边际主义观点就蕴含着其中任何人都会带来任何利益。假定在

救援情形中，一百个矿工被困地下，洪水正在上涨。如果四个人参与这个救援活动，这些矿工就会全都得救。

为了弄清其因果关系，我们可以假定，如果四个人站在台架上，合起来的重量就足以把每一个矿工升上地面。按照边际主义观点，如果有五个人参加这个使命，那么没有人会救任何人的命。这对于五个人的每个人都为真：这个人若不参与这个活动，不站在这个台架上，不会产生任何差别，因为其他四个人会救下所有这一百人的命。按照边际主义的观点，这些人中没有人做过任何好事。

这个结论看来是荒谬的。如果这些人没救人，那么这一百人的命是谁救的呢？有些作者主张，为了避免这种愚蠢的结论，我们应该诉求人们一起做事的效果。按照一种这样的观点，我们可以称为

总量平分观：当一组人一起产生了某个好效果时，每个人所产生的利益是总体利益的平均数。

254 这种观点蕴含着，如果五个人参与我们的救援活动，由此一起救了一百人的命，那么每个人应算作救了二十个人。任何一个人如果没有参与救援都不会引起差别，这是不相关的。基于这种观点，在判定何种可能的行动会产生最多的利益时，我们应该忽略每一个行动在被单独考虑的情况下的效果。

当讨论我们不偷盗和尊重别人财产权的义务时，休谟坚持了一种类似但更含糊的观点。休谟主张，正义和忠诚“对于人类的福祉是绝对必要的”。但是正

义的收益“并不是每一个单个行动的后果”，因为每个具体的正义行动，当“就其自身来考虑”时可能具有“极为有害”的后果。正义的收益仅来自于“整体框架”或“对普遍规则的遵守”。休谟因此主张，为了产生这些利益，我们必须遵循严格的规则，即使违背规则“就其自身来考虑”具有好的效果，也不允许有例外。这样的规则必须是严格的、不容变更的，因为“不可能把善与恶分开”。

基于休谟的观点，我们可以称之为

整体框架观 (Whole Scheme View)：要决定某个行动会产生多少利益，我们不应该问这个行动独自会产生多大影响。如果某个行动是一组行动中的一个，而这组行动一起会产生最大利益，那么该组行动中的每一个就都产生了这个最大利益。

行动后果论者如果拒绝边际主义观点而接受整体框架观，就可能会接受休谟的主张：我们应该遵循某些严格的规则，诸如“永不偷盗”，因为他们可能会认为，我们的每个行动就是以如此方式来产生最大利益的。这些行动后果论者由此也会是规则后果论者。如果整体框架观为真，这个主张就会是：

(C) 正是由于努力遵循最优规则，每个人才最有可能使得事情进展最好。

基于这些假设，这两种形式的后果论不会相冲突而是相一致的。

当康德捍卫另一个严格规则“从不撒谎”时，他提出了类似的主张。在一篇名声不佳的论文中，人们即使在面对杀人犯询问受害人之处时撒谎，康德也会为之做出谴责。这也导致人们常常假定，在主张我们绝不撒谎时，康德陈述的观点不可能是行动后果论。情况并非如此。康德写道，在讲谎话时，

255

我会尽力产生它，这个陈述……一般是不被人所相信的，这样基于契约的一切权利就会化为泡影，失去其效力，这是对一般而言的人性所造成的不公。

他还写道

由此一个谎言……总是伤害另一方，即便不是伤害另一个人也仍然伤害一般的人性，因为它使得权利的来源不可用。

在这些段落中，康德通过诉诸这样的行动所引起的伤害来谴责一切谎言。如前所述，拒绝边际主义观点且接受整体框架观的行动后果论也会做出这样的主张。康德可能像休谟一样相信，如果我们始终遵循某些严格的规则，我们每个行动就会产生最大的利益。

接着我们回到康德的主张：每个人应得的幸福会通过“最严格地遵守道德法则”而得到最好的促进。康德经常做出这样的主张。例如，他写道：

促进他人的幸福是一个目的，我能够提供的手段只能是自我完善，别无他法。

康德所称的“自我完善”主要在于我们具有善良意志和正当地行动。因此，康德在此的主张就是：正当地行动是促进他人幸福的唯一方式——或者说他可能意指的最好方式。康德还主张：

如果至善存在，那么幸福与配享由此就必定是相结合的。那么这种配享在于什么呢？在于行动与普遍幸福观念的实践一致。我们如果以这样一种方式——如果其他每个人都如此方式为人处事，就会产生最大的幸福——为人处事，那么这样做就应该配享幸福。

256 康德在此主张，要成为有德的和行动得当，我们必须这样来做事：如果每个人都采取同样的做法就会产生普遍的幸福。这种主张陈述了一种版本的后果论理论：享乐主义的规则效用主义。如果整体框架观和（C）为真，那么康德的观点也陈述了一种版本的享乐主义行动效用主义，因为这些观点是一致的。

然而，这些主张只具有历史的重要性，因为我们既要拒绝整体框架观，也要拒绝（C）。再次假定，在

救援情形中，一百个矿工被困地下，洪水正在上涨。如果四个人参与这个救援活动，这些矿工就会全都得救。我知道其他四个人已经参与了这个活动。我能够要么也加入这个活动，要么到别处救另外一

个人。

按照整体框架观，我应该加入这个活动，因为我的行动会因此是一组行动中的一个，而该组行动会是产生最多好处的行动，能救一百个人。这显然是个错误的结论。我应该去救那一个人，因为由此我们就可多救一人。至少在大多数情形中，我们应该接受边际主义观点。当我们问哪个行动会产生最多好处时，我们应该问这个行动会产生什么样的差别。既然我们应该接受边际主义观点，我们就不可能是行动-规则后果论者。后果论者就不得不在其观点的这两种形式之间做选择。

按照我们所称的康德的

最伟大善公式：每个人应该总是尽力促进一个普遍德性和配得幸福的世界。

正如我所论证过的，康德似乎认为：

K6：正是通过遵循如康德的其他公式所描述的道德法则，每个人才能最好地促进这个世界。

基于这些假设，康德的道德理论具有一种统一性或和谐，而这是康德宣称
为纯粹理性的目标之一。康德的最伟大善公式描述了一个单一的终极目的或目标，这是每个人应该总是努力去实现的；而康德的其他公式描述了道德法则，每个人遵循它就能最好地实现这个目标。

257

在确定我们是否应该接受些主张时，我们会有两个问题：

Q1：我们应该总是尽力去促进一个普遍德性且有配得幸福的世界吗？

Q2：我们通过遵循康德的其他公式能够最好地促进这个理想世界吗？

既然我们还没有考虑康德的其他主要公式（普遍法则公式）蕴含什么，我们就不可能试图回答Q2。

尽管我们可以试图回答Q1，但我不会那样做。然而，我将讨论康德有关他的理想世界的假设之一。人们有时候说，康德有关最伟大善的主张并没

给他的其余道德理论增添任何东西。康德在别处主张我们有两种目的，也是两种义务：我们自己的德性和他人的幸福。但是在描述他的理想世界时，康德补充了仅当幸福是应得的时候，它才是善的。基于康德的观点，如果人们具有比其应得的幸福更多或痛苦更少，那就是不好的。这些有关应得的主张无法可行地派生于康德的其他公式，或声称得到它们的支持。康德也没有尝试以这种方式支持这些主张。他只不过是坚持这些主张，或把它们看作是显而易见的，正如当他写道：

理性并不赞成幸福……除非在这样的范围内才可例外，即它一致于对幸福的配享，亦即道德的行动。

我认为，康德有关应得的主张是错误的。正如我现在会论证的，康德接近于明白这一点。

第十一章 自由意志与应得

38 道德所要求的自由

按照决定论，一切事件在因果上都是不可避免的，这样，我们只要以某种方式行动，那么在因果上就不可能以其他方式行动。康德主张，决定论若为真，我们就不会具有道德所要求的那种自由，道德因此就会受到破坏。康德认为，决定论以某种方式为真。但他主张，决定论并不是全部真理。康德区分了现象的时空世界与本体世界；现象是实体向我们显现出的样子，本体是实体如其真实所是地呈现的物自身。康德论证，在这个本体世界，既没有时间也没有空间。可以设想，我们除了是时空世界中的现象式存在外，还在这个本体世界中是本体式存在。我们的行动尽管在部分程度上是时空界里发生于时间里的事件，但在无时间的本体界可能具有不被决定的起源。康德认为，这会给予道德所要求的那种自由。 258

康德还论证我们具有这种自由。康德的论证可以陈述如下：

(A) 除非我们应当做出不同的行动，否则我们的行动就不可能是不当的。

(B) “应当”蕴含着“能够”。只有我们能做出不同的行动，我们才应当做出不同的行动。

因此

(C) 除非我们能做出不同的行动，否则我们的行动就不可能是

不当的。

(D) 如果我们的行动纯粹是时空世界里的事件，这些行动就是因果上被决定的，这样情况就不可能是我们能作出不同的行动。因此

(E) 如果我们行动纯粹是这种事件，我们就没有任何行动是不当的，这样道德就会是种幻觉。

(F) 道德不是一种幻觉。我们应当以某种方式行动，且我们某些行动是不当的。因此

(G) 我们的行动并不纯粹是时空世界里的事件。

在考虑这个论证时，我们首先也许会反驳，如果(E)为真，那么我们除非知道(G)为真，否则就不会知道(F)为真。如果除非我们的行动并不纯粹是时空世界里的事件，否则道德就是一种幻觉，且我们不知道我们的行动是否纯粹是这样的事件，那么我们又如何能知道我们的道德不是一种幻觉呢？但可能存在着某些方式，由此方式我们无需先知道(G)为真就能合理地认为道德不是一种幻觉。例如，这种信念可能为某组宗教信念所蕴含，而后者作为启示的真理，是我们能理性地接受且宣称知晓的。

我们还应该接受康德对(C)的论证。正如康德的假定，“应当”蕴含着“能够”。如果我们不可能以某种方式行动（诸如跑得比猎豹快去救人），那么我们应当以这种方式行动就不可能为真。我们的某个行动若不当，则我们本就应该采取另一种行动，因此“我们能够采取另一种行动”就必须为真。然而，“必须为真”的意思是什么，对此有着两种相冲突的观点，它们是有关道德要求何种自由的。

260 假定打雷时我正站在田地里，一束闪电从身边溜过。如果我说，我本来能被杀掉，我是在绝对意义上使用“能”的。我的意思可能是，即使是在真实发生的条件下，这束闪电击中我本来在因果上也是可能的。如果我们假定决定论，这就不会为真，因为这个闪电在因果上不可避免地会如它所做的那样击中地面。反过来我是在一种不同的、假设或虚拟的意义上使用“能”的。当我说本来能被杀死，我的意思可能只是，如果条件稍有不同（例如，如果我站得更靠西几步），我就会被杀死。即使我们假定决定论，这个主张还是能为真。

康德假定，仅当我们在绝对意义上做出不同行动时，我们才应当做出不同的行动。由此，必须为真的是如下情况在因果上是可能的：纵然我们实际的心态是既定的，我们本来也可以选择并实施其他行动。如果我们已做出的选择

与行动都是因果地不可避免的，那么我们本来能做出其他行动就不会在相关的意义上为真。基于这种观点，正如（E）所主张，决定论与道德所要求的那类自由是不相容的。

然而，正如许多作者所论证的，我们应当拒绝这种不相容主义的观点。回到这个情形，其中我说，“你本该帮助那个盲人过街”，而你说，“我本不可能那样做”。如果我问“为什么不能呢？”。而你的回答若是“因为我不想那样做”就是不够的。也许你在相关的意义上本不能做出不同的行动，如果你为某种不可控制的欲望所支配，或者是疯掉了。但我们大部分人的不自由都不是在如此种种的意义上说的。在大部分情形中，“我们能做出不同的行动”这点要在相关意义上为真，它只需要这点为真：

（H）我们若本来想或本来选择这样做，就会做出不同的行动。

我们可称此为假设性的、动机意义上的“能够”。这种意义上的“能够”与决定论是相容的。你本来能帮助这个盲人过街，意思是如果你本来选择了这样做，你就能这样做。如下情况是不相关的：鉴于你的实际欲望和其他心理状态，“你没有选择这么做”这一点是因果地不可避免的。

261

有些人可能反驳：

如果我们所有的决定、选择和行动都是因果上不可避免的，那么我们仅当奇迹般地违抗了或破坏了自然法则，才能做出不同的行动。我们是否应当以要求这种奇迹的某种方式行动，这种询问是无意义的。

然而，这种问题是很值得去问的。我们做什么，通常取决于我们应当做什么的信念。如果我们相信我们的某个行动是不当的或不合理的，因为我们本该做不同的行动，这个信念就可能导致我们改变自己或自己的情境，这样我们就不会再以同样的方式行动不当或不理性。我们或我们情境的这种改变可能会影响我们后面的所作所为。我们在过去要想做不同的行动，就得施行某个奇迹，这并不重要。如果我们变得相信我们应当做不同的行动，我们信念上的变化就可能使得下述为真：在类似的情形中，不需要任何奇迹，我们在将来就能做出不同的行动。这就足以使得询问“我们是否应当做不同的行动”是有价值的。

康德称此种相容主义观点为“拙劣的托辞”。他宣称，按照这种观点，我

们拥有的只是“转叉狗的自由”，即某种一旦上足发条就完全自动运转的机械装置。但康德对相容主义的反驳看来部分地取决于他没有做出另一个区分。

262 按照宿命论，我们随后会不可避免地以某种方式行动，无论我们决定做什么。我们一切不同的可能决定，都不过是我们终究会做同一件事的不同方式。按照这种观点，我们试图做好的决定是没有意义的，因为这丝毫不会影响到我们后面做什么。既然非常清楚，我们大部分人的行动确实取决于我们的决定，那么仅当宿命论限于某些特定的行动时，它才是可信的。例如，按照古希腊神话，无论俄狄浦斯决定做什么，他命定要杀父娶母。这点要为真，某个希借神不得不好做好干涉的准备，从而确保俄狄浦斯的决定无法防止他以后做这两件事情。

决定论是种极为不同的观点。按照这种观点，我们将来会做什么，取决于我们决定做什么。尽管我们的决定会是因果上不可避免的，但我们通常事先并不知道，并且也不可能总是知道，我们将来会决定做什么。如果我们做出更好的决定且据此行动，事情就有可能变得更好。这些事实足以给予我们理由努力做出好决定。如果我们认为努力做出好决定于事无补，我们就会错误地滑向宿命论，即假定我们的决定对于未来发生的事情毫无影响。

康德有时候犯了这种错误，如当他写道：

除非我们认为我们的意志是自由的，否则这种命令是不可能且荒谬的，留给我们的就只有等待并观察上帝会通过自然原因在我们之中注入何种决定，而不是我们自己作为创作者能做什么和应当做什么。

这些评论蕴含着，如果决定论为真，那么我们努力决定要做什么就全无意义了。我们就不得不是被动的，等着看我们会被导致做出何种决定。情况并非如此。即使决定论为真，我们也可以主动的，通过努力做出好的决定并据之行动。例如，如果我们正处于起火的建筑物中，我们可能努力决定如何逃生。如果我们只是干等着，瞧我们后来会被引起做何决定，我们就很可能做出更差的决定，更有可能死掉。

康德还提出了一种不同的相容主义的观点。他写道：

自由的实践概念与思辨概念毫无关联……因为我可能对于我现在的行动状态的源起完全漠不关心，我只问我现在不得不做什么，于是自由就是一种必然的实践命题。

康德在此似乎明白，当我们决定做什么时，我们能够忽略“决定论是否为真”这个思辨或理论问题。如果我们还不知道我们将决定做什么，我们在这个意义上就是自由的，即如果我们决定做什么，就没有任何东西阻止我们这样做。对于实践目的而言，这种相容主义的自由就是我们需要的一切。给定我们的实际心态，某个其他的决定是否在因果上不可能，这是不相关的。

尽管康德有时候表示，对于实践目的，我们所需要的自由与决定论是相容的，但他的主要观点显然是不相容主义的。康德甚至主张，本体上的无原因的自由是他整个哲学的基石。如果他接受的是这种相容主义的观点，他就不会提出这种主张。

按照我们一直在讨论的但更简要陈述的论证：

(A) 到 (E)：如果我们的行动纯粹是时间中的事件，这些行动就是因果上被决定的，并且道德就会是一种幻觉，因为我们没有道德所要求的那种自由。

(F) 道德不是一种幻觉。

因此

(G) 我们的行动并不纯粹是时间中的事件。

我已经主张，我们应当拒绝概括于 (A) 到 (E) 的推理。如果“我们的某个行动是不当的，则我们应当做出不同的行动”，那么“我们本来能做出不同的行动”就必然为真。但是相关意义上的“能够”是假设性的动机意义上的。这种意义上的“能够”与决定论是相容的。即使我们的行动是因果上被决定的，我们还是能具有道德所要求的那种自由。

39 为什么我们不可能应得痛苦

然而，还有另一种相容主义，康德正确地拒绝了它。康德的有些评论提出了这个论证：

(I) “我们的某个行动是不当的”要为真，我们就必须以某种可能使我们应得痛苦的方式为该不当的行动担负道德责任。

(J) 如果我们的行动纯粹是时间中的事件，我们就永远也不可能以这种应得痛苦的方式对这些行动负有责任。

因此

(E) 如果我们的行动纯粹是时间中的事件，我们就没有任何行动是不当的，这样道德就会是种幻觉。

因此

(G) 我们的行动不纯粹是时间中的事件。

前提(I)可能看来是可行的。人们认为，有些人无法以“使他们应得痛苦”的方式对其行动负有道德责任。例如，小孩子和疯子就属于这样的人。除了认为这些人无法以这种方式对其行动负责，我们还可能认为，出于这个理由，他们不可能行动不当。

这些人为什么不可能行动不当，还有一种更好的解释方式。小孩子和疯子不可能具有“行动不当”这一信念，也无法据此而行动。但是普通的正常成年人是可以这样做的。这就足以证成我们的信念：大部分人都是道德行动者，其行动可以是正当或不当的。这样我们应当拒绝康德的假设：我们要成为道德行动者，我们必须能以“使我们应得痛苦”的方式对自己的行动负责。我们能够融贯地认为：我们的行动既能是正当的或不当的，也没有任何人应得痛苦。

按照前提(J)，如果我们的行动纯粹是时间中的事件，我们就无法以这种应得痛苦的方式对我们的行动负责。我认为，康德的这部分观点是深刻的真理。我们能以几种其他方式或意义在道德上负责，但我认为，任何人都不曾能以可能使他们应得痛苦的任何方式负责。我也不认为，有任何人应得更少的幸福。

在康德对于假定(J)的理由中，有个信念是这样的：

(K) 如果我们的行动纯粹是时间中的事件，这些行动就会是因

果上被决定的。

以及

(L) 如果我们的行动是因果上被决定的，我们就从来不曾能以应得痛苦的方式对我们的行动负责。

我已经主张，道德要求的这类自由与决定论是相容的。若除了我们的欲望和其他的动机之外，没有任何东西会阻止我们做出不同的行动，则我们本来能在相关的意义上做出不同的行动。然而，正如康德所假定的，这类自由不足以证成这个信念：我们由于自己的所作所为而应得痛苦。康德这里正确地拒绝了我们可称为有有关应得的相容论。

在拒绝这种观点的其他人中，有些人可能会拒绝康德的这个主张：如果我们的行动纯粹是时间中的事件，这些行动就会是因果上被决定的。大多数物理学家现在不认为决定论为真，因为有些事件涉及到亚原子粒子，在部分程度上是没有原因或随机的。这样的主张并不适用于我们的行动决定或其他的心理事件。大部分神经学家认为，心理事件在因果上取决于我们大脑中的物理事件，后者是完全被决定的，因为它发生在太宏大的规模上，不会受到在亚原子层次的随机影响。但有些人拒绝这种观点，认为我们的有些决定不是完全在因果上被决定的。对于具有这样信念的来人，有些会诉诸亚原子层次的随机性。其他有些人是互动论的二元论者，他们认为心理事件并不是要么在于，要么在因果上完全取决于大脑中的物理事件。

要证成我们能应得痛苦这个信念，只捍卫“人们以某种方式行动的决定并不完全是被引起的”并不够。如果这就是我们对于任何这类决定所主张的一切，那么以康德的话来说，这就会是

相当于让它受控于盲目的机会。

基于这种观点，我们具有的不是转叉狗的自由，而是亚原子粒子的自由；前者的运动是因果上不可避免的，后者的运动则是随机的。当且因为我们大脑中的物质以某种随机方式运动或改变，我们因此不可能应得痛苦。如果正如某些二元论者所主张的那样，我们的决定在部分程度上是随机的非物理事件，这

也于事无补。

许多人主张，尽管大多数事件必须要么是完全被引起的，要么是部分程度上随机的，但我们的决定和行动不是这样的。这些人试图描述第三种可能性。其中有些人诉诸我们的理性。当我们主张某个人出于某个理由而行动时，这些人表示，我们不是主张这个人的行动是完全被引起的，我们也没有主张这个人的行动在部分程度上是随机的。我们出于理由而行动的能力看来提供了第三种选项。

然而，当某人出于某个理由而行动时，我们能问这个人为什么会出于这个理由而行动。在某些情形中，答案是由进一步的理由所提供的。例如，我撒谎的理由可能是掩盖我的身份，而我掩盖身份的理由可能是为了避免被指控某种罪行。但是我们很快就会达到任何这种动机性理由链条的开端。当我们达到某人以某种方式行动的终极理由时，我们能问这个人为什么出于这种理由而行动，而不是出于其他理由以其他方式行动。如果我有免遭惩罚的自利理由，又有不撒谎的道德理由，那么为什么其中之一对我具有更大的权重由此让我选择如此做呢？这个事件并不是因为某个进一步的动机性理由而发生的。这样一来，所提出的第三种选项在此就消失了。这个事件要么是因果上被引起的，要么是部分地随机的。在任何动机性理由的链条开端总是会有这么一个事件。既然我们实际上所做的所有行动决定都包含这样的事件，这里就不存在融贯的第三种选项。

267 为了避免这种争论，有些人主张，在不涉及任何事件的意义，行动可以由行动者引发。有些相信行动者因果性（agent-causation）的人部分地接受康德的观点：如果我们的行动纯粹是时间中的事件，那么我们就没有任何类型的自由使此为真：它使得我们能因为自己的作为而应得痛苦。但这些作者认为，作为行动者，我们完全是时空世界中的一个部分，这样他们无法合理地主张由行动者所引起的行动不是事件。

康德提出了其他相关的主张。康德假定，要对我们的行动负责，我们必须对我们自己的性格负责。用他的话来说：

人必须使得或已经使得他自己成为现在的样子……不管在道德意义上，是善或恶。这两种状况都必须是他自由选择的效果。

康德写道：

一个人的性格，这是他本人所创造的，

以及

一个人 (person) 是其自己的发起者 (originator)

亚里士多德类似地写道：

一个不正义的或自我放纵的人不变成那个样子，这在一开始是可能的，因此他们是自愿地成为那个样子的：但当他们变成这个样子后，他们要不是那个样子就不再可能了。

但亚里士多德并没有问，当一个人接近于使自己变成不正义和自我放纵的时候，“在一开始”本来能发生什么。康德问了这个问题，并且正确地主张，如果我们纯粹是时空世界中的存在者，我们就不可能自由地创造我们自己的性格，因此不能自由地选择成为善的或恶的。

随着刚才所引的这个主张，以及其他类似的主张，康德对于我们的行动不纯粹是时间中的事件这个信念提出了另一种论证。这个论证在部分程度上是

(M) 我们决定做的事情取决于我们的性格以及某些其他的事情，有关我们是什么样子的事实。

因此

268

(N) 要以“使得我们能应得痛苦”的方式对我们的行动负责，我们必须对“我们在相关方面是什么样子”负责。

(O) 如果我们的行动纯粹是时间中的事件，我们就不可能对我们变成什么样子负有责任，除非我们早前的行动方式使得我们成为那个样子。

(P) 要对早前的这些行动负责，我们必须对我们那个时候的样子负责，即通过更早前的行动使得我们变成那个时候的样子。

为了对这些更早期的行动负责，我们必须对我们早前那个时候的样子负责，即通过更早期的行动使得我们变成那个时候的样子。

为了对这些更早期的行动负责等……诸如此般直到无穷

(Q) 我们不可能对这种无穷系列的形成性格的行动负责。

因此

(J) 如果我们的行动纯粹是时间中的事件，我们就不可能选择我们自己的性格，或者以这种应得痛苦的方式对这些行动负有责任。

康德这部分论证是有效的，我认为它也具有真的前提，因此我们应当接受(J)。

康德的论证继续：

(R) 我们是以一种“能使我们应得痛苦”的方式对我们的行动负责的。

因此

269

(S) 我们的行动不纯粹是时间中的事件。我们对我们的行动负责，因为在无时间的本体世界中，我们自由地选择给予我们自己性格，去做我们所做的事情。

当其他作者试图描述除了是“充分被引起的或部分地随机的”之外的第三种选项时，对此的一个决定性反驳是“这种选项是无法理解的”。相较于这些主张而言，康德诉诸无时间的本体自由就某一方面而言更容易捍卫。康德主张，我们不应期待理解这种无时间的本体世界。我们能指望理解的是时空下的现象世界。用康德的话来说，尽管这种本体性自由是无法理解，但我们至少“能够理解其不可理解性”。

然而，我并不认为这是对康德观点的一个充分辩护。我们能含糊地理解某种实体可以是无时间性的。我们也能对这种观念有所理解：在某种非时间性的意义上，时空世界的所有特征都取决于模糊地类似于决定的东西。这种主张应用于上帝时可能有些意义。但是康德有关无时间性自由的某些主张甚至在这种

模糊意义上也是不可理解的。例如，基于康德的观点，尽管发生在时空世界中的每一件事都是因果上被决定的，但发生的每件事也是由无数自由且分立的决定所联合产生的，这种决定是由所有曾活过的理性存在者无时间性地做出的。我们无法想象，这样多的自由决定，其中有好有坏，能够挑选且产生整个完全被决定的事件系列，构成时空世界的整个历史。既然这些决定会部分地确定哪些理性存在者曾经存在，这些存在者就必定以某种方式展示其自身的存在。只说“我们至少能理解为什么这种主张是不可理解的”，这是不够的。我们能够理解这种主张没有可能为真。

按照我们正在讨论的这个论证：

(J) 如果我们的行动纯粹是时间中的事件，我们永远不可能应得痛苦。

(R) 我们能应得痛苦。

因此

(S) 我们的行动并不纯粹是时间中的事件。

我主张，我们应当拒绝这个论证的结论。我们的行动纯粹是时间中的事件。既然这个论证是有效的，且我们应当拒绝这个论证的结论，所以我们必须拒绝其前提之一。

有些人会拒绝(J)。也有人认为，尽管我们的不当行动纯粹是时间中的事件，是因果上不可避免的，我们能够应得被上帝送去地狱受苦。按照这种观点，为应得痛苦，我们没必要具有任何反因果的自由，或以任何方式对我们自己的性格负责，或对我们是谁负责。

在那些如此主张的人之中，有些人坦承，他们不能理解这种主张如何能为真。这些人主张，上帝的正义是不可理解的。相较于康德的这个主张（我们不应该指望理解无时间性的本体世界）而言，我们不应该指望理解我们能应得痛苦这个主张更说不通。我们没有任何理由预计这种道德事实是不可理解的。

我认为，我们应当拒绝的不是(J)，而是(R)。康德正确地主张

(J) 如果我们的行动纯粹是时间中的事件，我们不可能应得痛苦。

我们能补上

(T) 我们的行动纯粹是时间中的事件。

因此，

(U) 我们不可能应得痛苦。

271 我认为，康德接近于看出 (U) 的真理。康德认为

(V) 我们不可能应得痛苦，如果我们的行动是因果上不可避免的，或者受盲目机会的控制，以及我们不对我们自己的性格负责。

如果我们的行动纯粹是时间中的事件，那么康德认为这些情形就会为真。如果康德失去了对我们本体自由的信念，变得相信我们所有的行动纯粹是时间中的事件，那么他就有可能继续相信 (V)，并得出我们不可能应得痛苦的结论。但我不能宣称，我知道康德会得出这个结论。康德有可能反过来不再相信 (V)，并断定，即使我们的行动是因果上不可避免的，或者是受制于盲目的机会的，我们对于我们目前的样子是不负有责任的，我们还是能应得痛苦。我只能希望康德继续相信 (V)，因此本可明白我们不可能应得痛苦。

那些认为我们能应得受苦的人中间，有些人会给出这个反证：

(W) 上帝使得某些人在地狱中受苦。

(X) 上帝是正义的。

因此

(R) 我们能应得受苦。

但我认为，我们并不知道 (W) 为真。如果我们相信一个正义的上帝，我们必须接受

(Y) 上帝使得犯错者在地狱受苦是正义之举，这种受苦如何是一种正义之举是我们理解不了的。

或

(Z) 上帝没有使任何人在地狱中受苦。

我认为，对这两个主张而言，我们更有理由接受(Z)。如果上帝并没有使得任何人在地狱受苦，这么多的人认为上帝确实这样做了，这是令人吃惊的。我们能理解这些人如何变得具有这种错误的信念，但我们不能理解一个正义的上帝如何能使任何人在地狱中受苦。

272

我们可能应得许多东西，诸如感激、赞扬和某种只是道德非难的责备。但从没有任何人应得痛苦。我认为，出于类似的理由，没有人应得较少的幸福。当人们对我们或他人不当时，我们的愤怒是有道理的。我们能有理由想要这些人理解他们行动的不当，即使他们这会使得他们对其所作所为感到非常不好。但这些理由就像我们想要人们为他们亡去的爱人感到悲伤的理由一样。我们不能有证成地对这些行事不当者怀有恶意，希望他们过得不好。我们也不能有证成地对他们不再具有好意，不再希望他们过得好。我们最多地有证成不再喜欢这些人，以道德上可接受的方式不再与他们有任何关联。

如果康德已经明白没有任何人应得痛苦或不那么幸福，他的理想世界依然会是同一个世界，其中我们都是有德且有福的。但他会改变他有关不那么理想世界的观点，因为他不再相信：如果某些人的痛苦比其应得的要少，幸福则要多，会是坏事。

尽管康德对其理想世界提出了各种各样的主张，但这些并不是康德道德理论最有价值的部分。许多其他论者主张，最伟大的两种善是德性和幸福。康德很少捍卫其这一假设：如果我们遵循他的其他公式，我们所做的事情会最好地促进他的理想世界。最有价值的是康德理论中并不以这种方式成为后果论的那个部分。我们已经考虑了康德的人性公式，以及“要把人作为目的来对待的”其他相关主张：我们必须以他们能合理地同意的方式对待他们，永远也不能把他们作为纯粹的手段来对待。我们现在转向康德对其最高原则的其他主要陈述：普遍法则公式。尽管许多人已经讨论过这个公式，但我认为，没有任何人完全明白如赫尔曼所称的“替代后果论推理的这个选项所具有的未经开发的理论能量与活力”。

第三部分

理 论

第十二章 普遍法则

40 不可能性公式

康德声称，行为的正当与否取决于我们的准则（maxim）。康德的准则通常意指我们的策略（policy）及其隐含的目标，其中有这样一些例子：“运用一切稳妥的手段来增进财富”、“睚眦必报”（Let no insult pass unavenged）、“只要于己有益就许假诺（make lying promise）”、“不给需要帮助者提供任何帮助”，以及“自爱（self-love）或追求自身幸福的准则”。275

根据康德的普遍法则公式（Formula of Universal Law）的一个版本，我们可称之为

不可能性公式（Impossibility Formula）：按照任何不能成为普遍法则的准则来行动是不当的。

有必要解释这一公式。在一个段落，康德提及一种作为“普遍允许的（permissive）法则”的准则。这或许表明康德的意思是

（A）任何一个准则，如果不能允许所有人都按照它来行动，那么奉行该准则就是不当的。

但康德从未诉诸（A）。且如我在一个注释中所说明的，（A）不会是一种有用的主张。

有些作者提出，康德的意思是

(B) 任何一个准则，如果不能在决定按照它来行动的意义上为所有人都接受，那么奉行该准则就是不当的。

276

根据这一提议，康德的公式将不可靠。如果(B)谴责按照这样的准则——所有人都接受该准则是不可设想的或逻辑上不可能的——的行动，那么这一公式将不能谴责大多数不当的行动。我们能轻易地设想或想象这样的世界：其中每个人都接受诸如“只要利于己，即可欺骗和强迫他人”之类的坏准则。这样的世界或许在因果的意义上(casually)是不可能的，因为有些好人在心理上不能接受这些坏准则。但是，有些坏人也在心理上不能接受某些好准则。因此，(B)如果诉诸这种因果意义上的不可能性，就会错误地谴责按照这些好准则的行动。我们或许可以诉诸某种其他的不可能性。但如上述评论所表明的，(B)是讲不通的。我们没有任何理由相信：准则的好坏与否以及按照它们来行动是否错误，取决于是否每个人都接受这些准则。

有些作者提出，康德的意思是

(C) 某一准则，如果每个人按照它来行动是不可能的，那么奉行该准则就是不当的。

这里“每个人”一词仅仅指适用于某准则的人。例如，“照料自己的孩子”仅适用于父母。

这一公式也不可靠，因为(C)谴责许多为道德所要求或允许的行动。存在许多不能为一些人所奉行的好准则，因为他们没有按此方式来行动的机会或能力。有些父母由于身在狱中或精神失常而不能照料他们的孩子。但是，照料自己的孩子不是不当的。为了避免这一反驳，(C)可能谴责按照任何如下准则的行动：既有机会又有能力按照该准则来行动的每个人，不可能都按照它来行动。但是，没有任何准则不能通过这一检验。而且(C)也是不可信的。我们得不到任何准则，因为我们没有任何理由相信：准则的好坏与否以及按照它们来行动是否不当，取决于每个人能否按照它们来行动。

277

有些作者提出，康德的意思是

(D) 某一准则，如果不可能让每一个按照它来行动的人在实现其目标的意义上都取得成功(successfully)，那么奉行该准则就是

不当的。

这一公式毫无改进。有许多准则，我们按照它们来行动是可允许的或好的，但不可能全都取得成功。例如：“成为一名医生或律师”、“领养一个孤儿”、“给慈善机构的捐助比一般人多”，以及“做使用防火通道或离开沉船的最后一人”。我们如果全都努力去实现这些目标，那么其中一些人会失败。(D)也是讲不通的。我们没有任何理由相信：如果我们按照某一准则来行动而不能全都成功，那么任何人奉行该准则就是不当的。我们知道有些努力会失败，但这种努力并非不当的。

我们一直在力图理解康德的这一主张，即按照不能成为普遍法则的准则来行动是不当的。(A)至(D)是诠释该主张的最直截了当的方式。但是，它们不仅要么无益，要么既不可靠也讲不通，而且也不是康德本人在运用其公式时所要求的主张。尽管康德所陈述的(stated)不可能性公式是

(E)按照任何不能成为普遍法则的准则来行动是不当的，

但康德真正的(actual)公式是

(F)按照任何这样的准则来行动是不当的：如果每个人都接受并奉行该准则，或者每个人都相信按照它来行动是可允许的，那将使得任何人按照该准则来行动都不可能取得成功。

这一公式可以帮助我们确定哪些行动是不当的吗？

先考察这一准则：“只要于己有益就杀害或伤害他人”。如赫尔曼(Herman)所指出的：如果我们全都接受并奉行该准则，那并不会使得如此的任何行动都不可能取得成功。因此，(F)没有谴责这样的行动。(F)也没有谴责自利性的强制(self-interested coercion)。如果我们全都在有益于己的情况下设法强制他人，那么其中有些会取得成功。

再看撒谎。赫尔曼写道，(F)

对于欺骗的准则似乎是足够的……在康德看来，普遍的欺骗将使语言从而欺骗变得不可能。

与此相似，科丝嘉（Korsgaard）写道：

撒谎由于会骗人，通常能有效地实现目标；但如果被普遍践行，撒谎就骗不了人……

但是，无人会按照“总是撒谎”的准则来行动。许多撒谎者是按照“于己有益时撒谎”的准则来行动的。仅当在一个由自利的撒谎者组成的世界里任何撒谎都不可能成功时，康德的公式才谴责这样的行动。这并非不可能。即使在这样的世界，告诉他人真相也常常符合我们的利益。而在欺骗他人符合我们的利益时，撒谎常常变得毫无意义，因为对方不会相信我们的谎言。因此，即使我们全都是自利的撒谎者，我们所讲的也有很多是真话。大多数人都明白这一事实。而我们不可能总是判断出他人所讲的哪些话是谎言，因而有些谎言会为人所信，并实现撒谎者的目标。

为了解释偷窃为什么是不当的，康德写道：

每个人的财产都可被取走如果是一条普遍规则，我的（mine）和你的（thine）就都失去了意义。因为我可以从另一人那里取走的任何东西，第三者也可以从我这里取走。

然而，如前所示，没有人会奉行“总是偷窃”的准则。许多偷窃者奉行的是“在有利于己的情况下偷窃”的准则。如果这一准则被普遍接受和奉行，那并不会产生一个如此行动决不会成功的世界。财产仍将存在，但不总是受到有效的保护。偷窃者有时会实现其目标。

在讨论“睚眦必报”的准则时，康德断言，这一准则如果是普遍的，就会“自相矛盾”而不能“自圆其说”（*harmonize with itself*）。但是，如果每个人都奉行该准则，那并不会使情况变成这样：无人能够成功地做到。情况甚至可能是这样的：每一伤害均遭报复，由此每个人都会成功。

我们已经认识到，康德真正的公式未能谴责许多显然极为不当的行动。该公式没有谴责自利性的杀人、伤害、强制、撒谎、偷窃。

上述失败或许表明，康德的公式没有谴责任何东西。但我们还必须考察康德的最佳例证：某人许假诺以便能借到他无意偿还的钱款。此人奉行的准则是

“在有利于己的情况下许假诺”。康德断言，如果每个人都接受这一准则并相信许假诺是可允许的，那将使得任何这样的许诺都不可能成功。用他的话来说：

每个人……可以随意做出无意遵守的诺言，这一法则的普遍性（universality）将使许诺变得……不可能，因为无人会相信许诺，而把所有这样的表白（expression）嗤之为徒劳的借口（vain pretense）。

罗尔斯提议，在评估这一断言的过程中，我们应该过段时间，等到每个人都接受许假诺者的准则所产生的效果得以充分的展示，再问什么是正确的。康德的如下断言似乎是正确的：在这样一个世界，无人能用许假诺的方法为己谋利。不仅这样的许诺无人置信，而且受道德激励的（morally motivated）、包含信任的许诺这种社会实践（social practice）也不再存在。由此康德的公式谴责上述虚假的许诺。我们可以认为，大多数这样的行动是不当的。

既然找到了一类为康德公式所谴责的不当行动，我们就可以问问这一公式是否可行。康德的公式在一定程度上是

280

（G）按照任何这样的准则来行动是不当的：如果每个人都认为如此行动是可允许的，那将使得如此的任何行动都不可能成功。

这一断言谴责这样的行动：其成功依赖于他人由于认为该行动不当而克制不做。而且（G）的谴责似乎出于某种适当的理由。我们可以提出，许假诺者的行动是不当的，因为如果每个人都相信它是可允许的，那将损害一种有价值的社会实践。

然而，康德的断言未限定于有价值的社会实践。例如，希特勒（Hitler）军队中的军人被要求宣誓无条件服从。康德谴责许假诺，其所用的断言是：如果每个人都认为许假诺的行动是可允许的，那么许诺的实践将是“徒劳的借口”或摆设。其中有些德国军人正确地认为，他们尽管已宣誓，但违抗不道德的命令仍是道德上可允许的。我们可以做出类似于康德的断言：如果所有这些军人都认为这样的违抗是可允许的，那么宣誓无条件服从的实践将成为徒劳的借口或摆设。康德的评论似乎蕴含着，这种违抗是不当的。但是，如康德本人所主张的，每个人都应当违抗不道德的命令。

作为对(G)的又一检验,我们可以设想,在二战期间有位非犹太的德国公民知道德籍犹太人正被搜捕并遭杀害,他成功地奉行这样的准则:“在可以挽救某些犹太人的生命的情况下,向警察撒谎”。接着设想:如果已知每个人都认为这样的谎言是可允许的,那将使得任何人用这种方式去挽救人们的生命变得不可能。德国警察将被要求去搜查每间屋子,对于任何人关于屋内没有犹太人的说法,他们都将置之不理。根据这些假设,(G)本来应谴责该公民挽救生命的行动。

281 康德曾主张,即使是在蓄意谋杀者询问其意图谋害的对象身在何处时,我们撒谎也是不当的。有鉴于此,康德或许会接受上述结论。但是,这种挽救生命的谎言显然是正当的。而一旦运用于上述事例,(G)就是讲不通的。就这种救人方式而言,如果每个人都认为它是可允许的,就将使它不可能成功;这绝不会是对它的反驳。

这一假想例,类似于康德的许假诺者情形。康德的许诺者之所以能实现目标,是因为可以指望有许多人——鉴于他们认为这种许诺是不当的——不许假诺。康德主张,如果已知每个人都认为许假诺没错,那将使得任何人都不可能成功地奉行许假诺者的准则。这一点如果为真,康德的公式就意味着,此人许假诺是不当的。将类似的主张运用于我的例子。那位德国公民之所以能达成其目标,是因为可以指望有许多人——鉴于他们认为这样撒谎是不当的——不向警察撒谎。我已经假定,如果已知每个人都认为这样撒谎没错,那将使得任何人都不可能成功地奉行这位公民的救人准则。这一点如果为真,康德的公式就错误地蕴含着:此人的救人撒谎是不当的。这两种行动的重要区别在于,它们意图实现的目标不同;而这一区别为(G)所忽视。

如此种种情况表明,(G)是不可接受的。(G)未能谴责几乎所有明显极其不当的行动,与此同时还谴责某些显然正当的行动。而(G)尽管正确地谴责许假诺,却是出于坏理由去谴责。

康德的公式在一定程度上还是

(H)按照任何这样的准则来行动是不当的:该准则如果被普遍接受并奉行,那将使得任何人都不可能成功地按照它来行动。

282 有些作者声称,这一公式谴责依据数种好准则的行动,比如“拒受贿赂”、“慷慨资助穷人”。如果这些准则被普遍奉行,那不久就将使得任何人都不可能

成功地奉行，因为将无人会行贿，也不会再有穷人。因此，康德的公式错误地蕴含着：拒贿和慷慨资助穷人都是不当的。

科丝嘉对这一反驳做出了一定程度的回应。她提出，人们在奉行资助穷人的准则时，其目标是消除贫困。如果所有的富人奉行该准则，那就可以消除贫困，从而使得后来任何人都不可能奉行该准则。但是，(H)不会错误地谴责这些人的行动，因为通过资助穷人，他们将实现其目标。

然而，这样的主张不适用于某些富人。他们在奉行“慷慨资助穷人”的准则时，其目标不在于消除贫困，而在于他们由于慷慨而受到尊敬。如果所有的富人都按此准则行动，那么其行动可以消除贫困，从而使得任何这样的人都不可能用达成其目标的方式来按此准则行动。由此(H)将错误地谴责他们的行动。这些人在大笔资助穷人时，其行动没有道德价值，但并非不当。

接下来再看看那些接受荣誉规则 (codes of honour) 的人，比如让俄罗斯诗人普希金 (Pushkin) 在雪中作生死决斗的规则。假设普希金已接受这一准则：“决斗以显示我的勇气，但一直要朝天射击”。如果所有决斗者都接受并奉行这一准则，那么决斗的践行就变得很荒唐，而且不会延存。这会使得普希金不可能用实现其目标的方式来奉行其准则，因此，(H)将谴责普希金对于该准则的奉行。在此(H)或许看上去给出了正确的答案，因为决斗是不当的。但是，(H)原本不会谴责对于“决斗以显示我的勇气，且射死对手”这一准则的奉行。而奉行这第二个准则会坏得多。如这一比较所示，(H)谴责普希金的行动，这原本是出于坏的理由。如果普希金的准则被普遍接受，决斗的践行将无法延存；这原本绝不应成为反驳普希金准则的理由。如前所示，康德的公式不当地忽视了这一问题：某种社会实践是否优良以及是否应当得到支持。

作为又一例证，我们可以想想这样的准则：“不生养孩子，以便有更多的时间和精力为人类的未来而工作”。如果每个人都奉行这一准则，那将使得任何人都不可能成功地奉行，因为人类将没有未来。因此，(H)错误地谴责这样的行动。

283

奥尼尔 (O'Neill) 提出了 (H) 的一个较弱 (weaker) 版本。她提议，康德的公式可以成为

(I) 按照任何这样的准则来行动是不当的：一些人对该准则的成功奉行，将阻碍其他一些人对该准则的成功奉行。

奥尼尔声称，这一公式谴责欺骗和强制，因为欺骗或强制他人的那些人由

此“必然使得其受害者不能按照这些施害者奉行的准则来行动”。但这一断言为假。就受骗或被强制的那些人而言，其中大多数可以欺骗或强制另外的一些人。奥尼尔还声称，我们在欺骗或强制人们的同时，就“削弱了他们的能动性（agency）”，从而“至少在一段时间之内”阻碍着他们，使他们按照与我们同样的方式去行动不能取得成功。但这一断言也为假。两个人可以同时欺骗对方。也可以存在同时进行的相互强制。两名摔跤者可以同时用力，以让对方着地。而我可以发出一种可信的威胁来强制你，与此同时你也可以发出另一种来强制我。敌对的国家之间可以运用核武器相互威慑而使对方不敢使用核武器，其情形就是如此。

奥尼尔可能回答说，要表明（I）谴责欺骗和强制，如下断言就足够了，即，某些欺骗者和强制者阻碍他们的某些受害者，使得后者不能欺骗或强制别人。这一更弱的断言为真。奥尼尔有类似的断言：我们如果按照“重伤他人”的准则行动，我们之中的一些人就会使得其受害者中的一些人丧失能力，从而阻止后者去重伤他人。因此，（I）谴责一些不当的行动。但是，（I）是出于坏理由去谴责这些行动。欺骗、强制、重伤他人的不当之处并不在于：由于依此方式行动，我们得以阻止其他一些人，使之不能成功地做出同样的事。

此外，（I）错误地谴责许多好的或道德上可允许的行动。有许多好准则或可允许准则的情况是这样的：如果一些人成功地按照它们行动，那将阻止其他的一些人同样地行动。如奥尼尔所指出的，（I）意味着：带着争胜的目标进行竞技游戏，这种行动是不当的。尽管一些英国中小学男生被告知要接受这种观点，但这似乎太苛刻。而按照“成为一名医生”的准则行动并不存在任何不当，尽管通过申请并被允许进某个医学学校，我们阻止了其他某个人的准入。或者斟酌这些准则：“去发现恐龙灭绝的肇因”、“与他人一起旅行时总是背负最重的行李”，以及“找到某个可与我一起过幸福生活的人”。努力做出某项发现或背负最重的行李并非不当，即使我们的成功会使得他人无法这样做，这依然不是不当的。与某个人幸福地生活在一起也并非不当，即使他是另外某个人可以与之幸福生活的唯一的那个人，情况还是如此。

科丝嘉提出了康德不可能性公式的又一版本。她指出，这一公式所禁止的是这样的行动，其成功“依赖于它们是独特的”。她补充说，这种检验“揭示出不公平”。但我认为这不正确。而且康德公式的这一版本不当地谴责许多可允许的行动。有些穷人通过搜检他人扔的垃圾来获取食物。这个方法必定是独特的，但不是不当的，也并非不公平。浪漫的诗人让自己体验独自一人呆在

荒野，这并非不当。使用最不拥挤的网球场、要付利息之前偿还信用卡上的债务、只买二手书、开令人意外的舞会，这些也都不是不当的，也并非不公平。

尽管我们还可能用其他的方式来诠释或修正康德的不可能性公式，但这些可能性不值得考虑。在我们已经考察的诠释和修正之中，没有一个包含好主意。按照那种甚至不可能是普遍法则的准则来行动是不当的，这一断言没有任何有益的含义。

41 自然法公式与道德信念公式

康德提出另一个好一些的公式。根据其最重要的陈述

普遍法则公式 (Formula of Universal Law)：按照这样的准则来行动是不当的：我们不可能意愿 (will) 该准则成为普遍法则。

285

康德提到，对于那些不能通过这一检验的准则，我们具有不予奉行的不严格义务。这种义务是不严格的，其含义是指：有时道德上允许我们按照这样的准则来行动。应该忽略这一提法，康德也常常这么做。康德声称，我们的严格义务可以从他的不可能性公式中派生出来。如上节所述，这并不正确。因此，我们应该问：借助正确地蕴含“某些种类的行动总是不当的”，康德的普遍法则公式能否有所改进。如赫尔曼所指出的，康德的公式如果有这样的蕴含——为一己之便而杀害他人的策略尽管是不当的，但有时是允许的——那就是不充分的。

在运用康德的公式时，我们设想或想象，我们有能力去意愿或选择“某些事情是正确的”。我们在做某种思想实验 (thought-experiment)，这种实验涉及比较世界的不同可能状态，或者可称之为不同的可能世界。像一些科学家的思想实验一样，关于这些可能世界的思想可以引导我们达至也适用于现实世界的结论。

在问我们能否意愿“某一准则成为普遍法则”为真时，康德有时问的是，我们能否一贯地 (consistently) 意愿其为真。例如他提问说，我们的意愿是否会与自身相冲突，或者会与自身相矛盾。在其他段落，康德似乎问的是，我们能够合理地 (rationally) 意愿或选择什么。如果我们是在较宽泛的、第二种意义上使用“能够意愿”，康德的公式更可能成功。在某些观点看来，这毫无影响，因为我们的选择仅当不一贯或相互冲突时才不是合理的。但如我已论证

的，我们的选择要成为合理的，还必须对理由或显见的（*apparent*）理由予以充分回应。我们不能合理地选择或意愿“某一准则成为普遍法则”为真，只要我们意识到，有些事实给予我们以显然为决定性的、不这么选择的理由。

就意愿某一准则成为普遍法则而论，我们会意愿什么？康德有时声称，在运用他的公式时应该问的是，我们能否意愿我们的准则在每个人都会接受和奉行的意义上是一种“普遍的自然法”。关于康德公式的这一版本，我们可称之为

自然法公式（*Law of Nature Formula*）：按照某准则行动是不当的，除非我们能够合理地意愿如下情况为真：每个人都接受并尽可能地奉行该准则。

如前所示，“每个人”一词仅指可以按照某准则来行动的人。例如，“戒烟”的准则仅适用于吸烟者。

在一些其他的段落，康德诉诸我们可称为的

可允许性公式（*Permissibility Formula*）：按照某准则行动是不当的，除非我们能够合理地意愿如下情况为真：在道德上允许每个人都奉行该准则。

在运用该公式时，康德认为，如果允许每个人都奉行某一准则，那么至少有一些人更有可能奉行。这种效果之产生，并不是由于这些人被允许奉行该准则，而是由于他们相信如此行动是被允许的。因此，康德必定也在诉诸我们可称为的

道德信念公式（*Moral Belief Formula*）：按照某准则行动是不当的，除非我们能够合理地意愿如下情况为真：每个人都相信奉行该准则在道德上是可允许的。

这两个公式既然相似就没有必要同时使用。而如我的一个注释所解释的，与可允许性公式不同，道德信念公式独自使用也能讲得通。因此，我们可以忽略前者。

康德谈到，他所提出的并非某个“新原理”，只不过是更准确地陈述了“普

通人类理性……总能显而易见的”原理。这个说法低估了康德的原创性。但康德的自然法公式和道德信念公式所提出的观点，可以用两个常见的问题来表达：

“如果每个人都那么做会怎样？”与“如果每个人都像你那么思考会怎样？”

在运用这些公式时，我们必须诉诸某些关于合理性和理由的信念。我们可以诉诸康德本人的信念。但这会有困难，因为康德对其信念未作清晰的说明。而我们要问的是，康德的公式能否有助于我们决定哪些行动是不当的，以及能否有助于解释这些行动为什么是不当的。在提问过程中，我们应该设法诉诸关于合理性和理由的正确信念。因而我们应该诉诸自己的信念，因为由此就是在诉诸自认为最正确的或最佳的观点。尽管知道可能弄错，但我们不可能不诉诸自认为正确的东西，而诉诸事实上是（is）正确的东西。

然而，有些信念是我们不应该诉诸的。首先，我们不应诉诸关于哪些行动是不当的信念。我将之称为我们的道义式信念（deontic beliefs）。我们也不应诉诸由某一行动之不当性（wrongness）可能提供的道义式理由（deontic reasons）。在运用康德的自然法公式时，同时声称如下的（1）和（2）两者，那将毫无意义：

（1）按照某一准则行动是不当的，因为我们不能合理地意愿“每个人都奉行这一准则”为真。

以及

（2）我们不能合理地意愿“每个人都奉行某一准则”为真，因为这样的行动是不当的。

如果将两者合并，那就会像试图扯着鞋带让自己悬空。换个譬喻，我们就在原地打转，毫无进展。康德没有犯这种错误。在声称我们不能合理地意愿“每个人都奉行某种坏准则”为真时，康德从未既诉诸他关于“如此行动是不当的”信念，又诉诸我们不能合理地意愿“每个人都行动不当”为真。康德知道，如果诉诸这样的信念，其自然法公式就将一无所成，因为该公式由此就不能帮助我们取得关于哪些行动为不当的正确信念，也不能支持这些信念。

类似的评论也适用于康德的道德信念公式。同时声称如下的（3）和（4）两者，那将毫无意义：

(3) 按照某一准则行动是不当的，因为我们不能合理地意愿“每个人都相信这样的行动是可允许的”为真。

以及

(4) 我们不能合理地意愿“每个人都相信这样的行动是可允许的”为真，因为这样的行动是不当的。

在问我们能否合理地意愿“每个人都相信某种行动是不当的”时，我们不应该诉诸这些行动为不当的信念。如前所示，在运用道德信念公式时，康德遵循了这一道义式信念的限制（*Deontic Beliefs Restriction*）而没有诉诸这样的信念。

还有另一种我们不应诉诸的信念。许多不当的行动有利于行动者，其获利方式是通过给他人带来大得多的负担。在某些观点看来，这样的行动是不合理的，因为给予其他每个人的福祉（*well-being*）以充分的权重，这是对我们的合理要求。我们如果接受这一观点，那么在运用康德的公式时应予以搁置。康德的自然法公式蕴含的主要观念在于，犯错者（*wrong-doer*）即使能够合理地按照某些坏准则来行动，但不能合理地意愿“每个人都按照他们的准则来行动”为真。在运用这一观念时，我们做出如下断言将是不相干的：给予他人的福祉以充分的权重是对这些人的合理要求，因而这些人甚至不能合理地意愿“他们本人按照其准则行动”为真。

如前所示，康德没有做出这样的断言。康德讨论了一个自立的（*self-reliant*）富人，他持有“不帮助需要帮助者”的准则，此时康德没有诉诸这样的信念，即给予帮助是对这个富人的合理要求。如罗尔斯和赫尔曼所提议的，在将康德的公式运用于奉行如此准则的人时，我们应该假定，这些人的准则和行动是合理的。我们可以补充说，如果将康德的公式与较少争议而较广为接受的、关于合理性与理由的假定相结合，那么这些公式只要成功就会取得更大的成效。

42 行动者的准则

康德的公式认为，某一行动是否错误，取决于行动者的准则。在康德讨论的准则之中，大多数包含某种可以奉行于好几种情形之中的策略。两条包含同样策略的准则或许是不同的，因为它们隐含的动机或目标不同。例如，两个商

人可以都奉行“绝不欺骗顾客”的策略。但是，如果其中一位是由于他认为这是其义务，而另一位的动机是维护其声誉和赢利，那么他们所奉行的准则就是不同的。

康德诉诸行动者的准则，这会引发各种问题。让我们称某一准则是

普遍的：只要每个人都尽可能奉行该准则且相信这样的行动是允许的。

假设我做了件错事：偷窃某个女人的钱包；这个女人身穿白衣，正边吃草莓，边读斯宾诺莎（Spinoza）《伦理学》（*Ethics*）的最后一页。我的准则是，只要有可能就精确地按照如此方式行动。我可以合理地意愿“这一准则是普遍的”为真，因为对于其他任何人而言，精确地按此方式行动的可能性是极小的。因此，“这一准则是普遍的”会产生影响的情况极其不可能。在奉行这样的准则时，犯错者能够合理地意愿其准则成为普遍的，因为他们知道，这样的行动将极少再现，从而几乎不会产生任何影响。康德的公式将错误地允许这样的不当行动。我们可称之为稀罕性反驳（Rarity Objection）。

这一反驳可以得到部分的答复。正如某人相信、想要或意图什么是一个事实性（factual）问题一样，某人按照哪种准则行动也是一个事实性问题。而现实中的人很少按照如此高度具体的准则来行动。如奥尼尔以及其他一些人所主张的，在描述某人的准则时，我们不应纳入任何这样的细节：无论这个人在做什么，该细节的缺乏都不会对他的决定产生任何影响。就我的例子的现实版而言，即使失窃者穿的是红衣服、是在吃蓝莓、或者是在读《对喔，吉福斯！》（*Right Ho Jeeves!*）的首页，我仍会偷窃。我真实的准则将类似于“在有利于己时偷窃”。这或许不是一条我能够合理地意愿其为普遍的准则。由此康德的公式就正确地蕴含着：我的行动是不当的。

290

这些意见没有完全答复稀罕性反驳。即使实际的犯错者从未按照如此高度具体的准则行动，我们也可以想象出这样的人存在。康德的公式应当有能力谴责这些想象中的人的行动。而且如我们将会看到的，这一反驳适用于某些实例。

康德诉诸行动者的准则，还会引发另一更严重的问题。考虑某个人通常奉行

利己主义准则（Egoistic Maxim）：做最有利于己的任何事。

我们可以有理由认为，这个人不能合理地意愿如下的任一情况为真：每个人总是都按此准则行动，或者每个人都相信所有这样的行动是道德上可允许的。大多数利己主义者不能合理地选择生活在一个由利己主义者组成的世界，因为对于他们来说，那将比生活在人们接受多种道德准则的世界要糟糕得多。由于这种利己主义者不能合理地意愿其准则成为普遍的，康德的公式就蕴含着：他只要按照其准则行动，其行动就是不当的。这个人不仅在偷窃和撒谎时的行动是不当的，而且在出于自利的理由去还债、守诺乃至救落水儿童时的行动也是不当的，因为他希望获得某种回报。这是不可接受的结论。这种利己主义者以此方式行动时，其行动没有道德价值。但是，这些行动并非不当。

291 或许可以声称，这个人在按照上述方式中的任一种来行动时，他在做的事（*what he is doing*）不是不当的，但他对此事的做法（*his doing of it*）是不当的。康德做出了一个类似的区分——他声称，要履行某些美德的义务（*duties of virtue*），我们不仅必须正当地行动，而且必须出于正当的动机而行动。罗尔斯断言，在康德看来，我们即使没有自杀，也可能未履行不自杀的义务。要履行这一义务，我们必须出于正当的理由而克制自杀。康德有类似的断言：要履行感恩的义务（*duty of gratitude*），我们必须怀有感恩之情。

这些区分不能答复对康德公式的上述反驳。例中的利己主义者也许从未履行其美德的义务，因为他也许从未持有正当的动机。然而如康德所声称的，我们还有许多正义的义务（*duties of justice*）——无论我们的动机是什么，我们都可以通过做道德上要求做的事来履行之。一个例子是还债的义务。康德所讲的精明商人如果按照“偿还我的债务”的准则行动，那就履行了这一义务，即使其唯一的动机是维护声誉和赢利。康德的公式在此给出了正确的答案，因为这个商人将会是按照他能够合理地意愿其为普遍的准则来行动。但是，我所讲的利己主义者在还债时是按照其利己主义准则行动的，他不能合理地意愿这一准则是普遍的。因此，康德的公式错误地蕴含着：这个人在还债时不是在履行其义务，而是在不当地行动。

再回到落水儿童的例子。假设这个孩子的落水之处是靠近深水瀑布的一个水流湍急的河流，因而救他的任何努力都非常危险，超出任何人的义务范围。某个好人如果不顾危险去救这个孩子，那将是超出义务要求的英雄之举。我所讲的利己主义者决定甘冒这些危险，因为由此他有望得到某种更大的回报。他按照其准则行动，跳入河中。按照我们现在考虑的提议，如果这个人冒着丧命的巨大危险去救这个孩子，那他在做的事没错，但他对此事的做法是不当的。

这显然不正确。此人没有疏于义务的履行，也没有任何意义上的行动不当。

继而转到不影响其他任何人的审慎（*prudent*）行动。在服药或加衣保暖时，这个利己主义者或许是在按照“做最有利于己的任何事”的准则行动。他不能意愿其准则成为普遍的，因而康德的公式又会不当地蕴含着，他的行动不当。我们也不能声称，尽管他所做的事没错，但他对此事的做法是不当的。如此说法毫无道理：他在加衣保暖时的行动在此意义上是不当的。

有些作者提出，我们不应将康德的公式运用于像“做最有利于己的任何事”那样的笼统（*general*）准则。但是，康德经常讨论这种利己主义准则，他称之为“自爱或自身幸福的准则”。而我们如果声称这样的准则过于笼统，那将忽视许多人的实际准则。康德讨论了“在有利于己的情况下许假诺”的准则。还有其他类似的准则，比如只要最有利于己就偷窃、欺骗或违法的准则。既然这些准则全都包含同一种更为一般的策略，那就没有必要这么零散，而全部可以由“做最有利于己的任何事”这单个准则所取代。现实中有不少人按照这种利己主义准则或策略来行动，在此情况下声称这些人也接受并奉行其他某种一般化程度较低的策略，这或许可直斥为虚妄。

我们可以转而考察不同类的例子——持有错误道德信念的有良知者。可为其中一例的是：如康德的某些评论所表明而我们在此可假定的，康德本人有一时期接受“从不撒谎”准则。这一准则为康德的公式所谴责。但康德本不可能合理地意愿“无人会撒谎，即便是在意图谋杀者问询其谋杀对象的处所时”为真，也不可能合理地意愿“每个人都相信那些救命的撒谎是不当的”为真。因此，康德的公式就蕴含着：只要康德按照该准则行动、对任何人都讲真话，那么他的行动就是不当的。甚至在告诉某人以某天的正确时辰时，他的行动也是不当的。这显然是错的。类似的断言将适用于接受“绝不偷窃”和“绝不违法”准则的人。这些人不能合理地意愿“无人会偷窃或违法”为真，即便在这样的行动是拯救某一无辜者生命的唯一途径时也是如此。因此，康德的公式就蕴含着：这些人只要通过偿还财物或守法而奉行这些准则，那就是在不当地行动。这些蕴含显然也是错的。

我们的问题可重述如下。某些准则在总是不当的意义上是全然坏的（*wholly bad*），或者在总是正当的意义上是全然好的（*wholly good*）。“折磨他人以供我取乐”与“阻止无谓的痛苦”即是其中的两个例子。在运用于这样的准则时，康德的公式有效。但是，许多准则是

道德混合型的 (morally mixed)，其意思是：如果总是按照这些准则行动，那么其中有些行动是不当的，但另一些行动是可允许的，甚或是为道德所要求的。

利己主义准则与康德“从不撒谎”的准则，即是其中的两个例子。在提出公式的过程中，康德忽视了这样的混合型准则。康德的公式假定，按照某一准则行动，要么总是不当的，要么绝非不当。在运用于混合型准则的情况下，康德的公式会失效，因为这些公式谴责某些显然可允许的或为道德所要求的行动。在利己主义者精明地还债、康德对大多数人讲真话的时候，他们并不是像康德的公式所错误地蕴含的那样在不当地行动。我们可称之为混合型准则反驳 (Mixed Maxims Objection)。

对于康德的普遍法则公式——无论是其自然法版本还是道德信念版本，在考察了形形色色的反驳之后，有些作者断定，不可能运用康德的公式来帮助我们确定哪些行动是不当的。伍德 (Wood) 声称，康德的公式在用作这样的标准时是“根本上有缺陷的”和“近乎无价值的”。赫尔曼声称，尽管有一段“令人遗憾的努力史……但没有任何人能够使之变得可行”。奥尼尔提出，在某些情况下，康德的公式可以“给出的，要么是不可接受的指导，要么是根本没有”。希尔 (Hill) 怀疑这一点：康德的公式在独自使用的情况下，能否提供“即便是某种不严谨的 (loose) 和局部的行为 (action) 指南”。

这些人认为，康德的公式不能提供关于不当性的标准。因此，其中有人提出，康德并没有力图提供这样的标准。赫尔曼指出，康德的公式或许仅仅旨在表明：存在着反对出于某些理由、按照某些方式行动的某种“审慎的推测” (deliberative presumption)。奥尼尔提出，康德的公式或许旨在提供一种检验，不是检验哪些行动是不当的，而是检验哪些行动具有道德价值。

294 我认为，康德有更具雄心的目标。用康德的话来说，在某种意义上，当我们的行动符合义务 (conform with duty) 或违反义务 (contrary to duty) 时，这些行动就是正当的或是不当的。这是康德的公式所关注的“正当的”与“不当的”的含义。在讨论或运用其公式时，康德写道：

用最简短但绝对可靠的方式告诉自己……一个虚假的承诺是否符合义务，我会自问：我的确会对我的准则满意吗……应当持之为一条普遍法则吗？

某人厌倦生命……但请自[问]，剥夺他自己的生命是否违反对他自己的义务。

他仍然有足够的良知自问，这不是被禁止和违反义务的吗？

他自问，怠慢其自然禀赋的准则是否……与人们所称的义务相一致。

康德还断言，他的公式

非常准确地确定什么是该做的……就所有的一般性义务而言，

以及

有此罗盘在手，普通的人类理性就十分清楚，在每种情形之下如何区分孰善孰恶，何者符合义务或违反义务。

后面的这些断言讲过头了。但我认为，如下的断言也讲过头了：作为一种检验不当的标准，康德的公式毫无价值，并且不可能变得可行。康德的公式能够变得可行。我将要论证的是：一旦以某些全然（wholly）康德主义的（Kantian）方法来修正，这一公式会取得惊人的成功。

要问应该如何修正康德公式的两个版本，我们可以先重申混合型准则反驳。要判断某一行动是否不当，我们必须知道与之实际或可能的道德相关的（morally relevant）所有事实。例如，知道这些是不够的：某人动了他的一根手指，或动这根手指的同时扣动了枪支的扳机，或他由此杀死了某人。我们还必须知道某些其他的事实。比如，这个人当时是否意图杀另一人；如果是，那他是否出于自卫而杀人；如果是，那他自卫时是否在袭击另外的某个人。

295

我已经讲过，在康德讨论的准则之中，大多数准则包含某种可以奉行于好几种情形之中的策略。康德的公式假定，要判断某人的行动是否不当，知道他奉行的策略就足够了。这有时是正确的。知道某人在奉行“折磨他人以供我取乐”的策略，这是足够的。但在其他许多情况下，康德的假定是失败的。对于例中的利己主义者，如果所知的一切只是他在奉行“做最有利于己的任何事”的策略，那我们就不能确定他是否行动不当。我们不知道，这个人是在杀人、救人、偷窃、还债，还是在加衣保暖。而如果所知的一切只是康德奉行

“从不撒谎”的策略，那我们就不知道他是让某个意图谋杀者知晓其谋杀对象之所在，还是仅仅告诉某人以某天的正确时辰。这些例子表明，我们如果所知的一切只是某人在奉行的策略，那通常就未能了解所有那些可能是道德上相关的事实。

还存在另一个问题。在问某一行动是否不当或违反义务时，康德的公式时常依赖于道德无关的（irrelevant）事实来作答。例中的利己主义者在冒生命危险去救落水儿童时，他在奉行“做最有利于己的任何事”的策略是无关的。康德在告诉某人正确时辰时，他在奉行“从不撒谎”的策略是无关的。这些事实至多给予我们理由去相信，在某些其他的情况下这个利己主义者和康德可能会行动不当。

康德的公式要取得成功，那情况就必须是：绝不存在这样的准则或策略——奉行它们有时但并非总是不当的。这显然是错误的。因此，在“准则”能够意指策略的意义上，康德的公式不应该诉诸行动者的准则。

296 有些作者提出，康德的公式不应诉诸行动者的实际准则，而应诉诸行动者本来可以奉行的可能准则。就其自然法（LN）版本而言，康德的公式由此就可以成为

LN2：我们在不当地行动，除非我们正在做的是，在我们按照某个能够合理地意愿每个人都奉行的准则来行动时原本可做（could have done）之事。

这个公式避免了混合型准则反驳。在例中的利己主义者救落水儿童、康德对大多数人讲真话的时候，他们原本可以按照他们能够合理地意愿每个人都奉行的准则来行动。但如果诉诸LN2，就要损失我们对稀罕性反驳的部分答复。回到我不当地偷窃穿白衣、吃草莓的女人的情形。我在做的是，在尽可能按照“偷窃穿白衣、吃草莓的女人”的准则行动时原本可做之事。我能够合理地意愿“每个人都奉行该准则”为真，因为如此行动即使还有也是极为罕见的。因此，LN2错误地允许我的行动。类似的断言适用于其他的例子。在人们不当地行动时，总会存在某种符合如下要求的可能准则：这些人能够合理地意愿该准则为普遍的，并且原本可以按照此准则行动。因此，LN2不能谴责任何不当的行动。

为避免这一反驳，我们可以用更简明的方式来修正康德的公式。康德的自

然法公式可以成为

LN3: 我们在不当地行动, 除非我们在做之事是, 我们能够合理地意愿每个人在类似情形下都会尽可能做的。

康德的道德信念公式 (MB) 可以成为

MB2: 我们在不当地行动, 除非我们能够合理地意愿“每个人都相信这样的行动是道德上可允许的”为真。

这两个公式避免了混合型准则反驳。在例中的利己主义者救落水儿童、康德告诉某人正确时辰的时候, 他们能够合理地意愿如下两种情况均为真: 每个人都按照这样的方式行动, 以及每个人都相信这样的行动是可允许的。因此, 这两个公式未犯不当地谴责这些行动的错误。

这两个经修正的公式也避免了稀罕性反驳。在将这些公式运用于某人的行动时, 我们必须以道德相关的方式来描述这个人的行动。假设我患有古怪的偷窃癖, 确实是在奉行“尽可能对穿白衣、吃草莓的女人实施偷窃”的准则。这一准则没有提供有关我的行动的道德相关描述。我实施偷窃的对象是一个女人、她身穿白衣、在吃草莓, 这些都是无关的。相关的事实或许在于, 我是在偷窃某个不比我富裕的人, 目的仅是取乐。我们在运用这些修正公式时应问一问, 我能否合理地意愿如下情况为真: 每个人都按照这种方式行动, 以及每个人都相信这样的行动是可允许的。如果答案如我们能合理地主张的那样是否定的, 那么经修正的公式就会正确地谴责我的行动。

297

在许多情况下, 要给出对某一行动的道德相关描述, 描述行动者正在或将要意图做 (intentionally doing) 什么即已足够。我们必须描述这个人的直接目标 (immediate aims), 或这个人力图直接实现的是什么。还应描述这个人相信他/她的行动可能产生的效果。人们意图做 (intentionally do) 什么, 不同于他们意图 (intend) 什么。以西季威克的举例看, 如果19世纪末某个俄罗斯革命党人炸掉沙皇 (Czar) 旅行乘坐的列车, 那么这个人或许只是意图杀死沙皇。但这个人正在意图做的是炸掉列车, 同时还知道这样不仅会杀死沙皇, 而且会杀死许多其他的人。

在描述人们的行动时, 我们通常是在描述这些人在意图做什么。有时不

清楚的是，对于某一行动的道德相关描述是什么。譬如说，或许不清楚的是：应当将多少事实纳入我们关于某一行动的可预见效果之目录，或者应当将何者描述为分立的行动（*separate acts*）、将何者描述为单个复杂行动的组成部分（*parts of a single complex act*）。而要确定某一行动是否错误，我们有时需要知道的，不仅是某人在意图做什么，而且包括这个人为什么做其在做之事。为说明这两点，我们可以设想，某个施虐狂救了某人一命，以便他随后能以更令人痛苦的方式杀死此人。此时声称这个施虐狂在意图做的是救某人的生命，这或许是不够的。

298 在不清楚某一事实是否为道德相关的情况下，将之纳入我们对行动的描述通常是无害的。但在将某些道德原则运用于某一行动时，可能很重要的是不纳入道德上无关的事实。我已经讲过，要同时运用LN3和MB2，我们必须对人们正在做什么给出正确的描述。类似的主张适用于某些其他的道德原则，比如关于撒谎、偷窃和违背诺言之不当性的原则。有时不清楚的是，应将哪些行动看做是属于这些类型。但我们在此不必答复这些问题。我的主要主张是，在众多情况下，行动者的准则没有给予我们关于某一行动的道德相关描述。

按照我建议的康德公式的版本，我们将不再使用康德的准则概念。或许有人提议，我们可以在较狭窄的意义上使用“准则”一词，使之不包含某人在奉行的策略，而仅指这个人在做什么。康德有时以此方式来使用“准则”，如他在讨论“自杀以免除痛苦”准则的时候。这个准则不是一种策略，因为我们仅能按此行动一次。但是，对于行动的道德相关描述而言，这种较狭义的“准则”不会有任何增益。

对于康德对准则概念的使用，我们在此可以再补充一个反驳。人们在行动时常常并没有奉行任何策略。我们如果使用“准则”一词仅指策略，那就不得不承认，存在众多无准则的（*maximless*）行动。要能涵盖这样的行动，康德的公式必须时常使用“准则”这个词去指涉的，不是某种策略，而是（根据对这个人的行动的道德相关描述）某人在做之事。由于康德的公式通常必须直接运用于人们的行动，这就难以说明，这些公式为什么应该总是意指人们的策略而非他们的行动。

或许有人提出反驳，如果以放弃准则概念的方式来修正康德的公式，我们就不再是在讨论康德的观点。这一断言是真的，但丝毫不构成反驳。我们在问的是，康德的公式能否有助于我们确定哪些行动是不当的，能否有助于解释这些行动为什么是不当的。我们如果能以显然必要的方式修正这些公式，那就是

在发展某种康德主义的道德理论。而我认为，康德对准则概念的使用，并非康德自身理论之中有价值的部分。就终止使用这个概念而论，我们没有损失任何值得坚持的东西。

或许有人会质疑上述断言。奥尼尔写道，康德诉诸行动者的准则，这并非“康德理论中可分离或不重要的部分”，因为康德观点的这一特征使我们得以宣称：某个犯错者意愿其坏准则成为普遍的，此时他的意愿之中存在矛盾。我们由此可以论证说，恶行（wrong-doing）包含着“不能持有融贯的（coherent）意图”。但如康德所指出的，犯错者事实上并不意愿其准则成为普遍的。因此，这些人的意愿之中“实际上毫无矛盾”。

奥尼尔还提出，通过诉诸行动者的准则，康德回答了“对人们行动的道德相关描述是什么”这一问题。但如我们所见以及奥尼尔在别处所主张的，事实并非如此。如果所知的一切只是例中的利己主义者奉行其准则，我们就不可能确定这个人的行动是否不当。

接下来或许还有人反对说，如果把康德的公式修改到甚至不涉及准则，我们就会损失康德观点之中另一个有价值的部分。康德将准则定义为行为的主观原则，并且问我们能否意愿这种原则是一种普遍的法则。我们的康德公式版本没有涉及原则或法则。但MB2可以重述为

MB3: 我们在不当地行动，除非我们能够合理地意愿“每个人都接受某种允许这样行动的道德原则”为真。

这个版本保持了康德对原则和道德法则的关注。

现在回到奥尼尔的提议：通过将康德的公式运用于行动者的准则，我们至少能够确定某一行动是否具有道德价值。该提议有一些道理，因为一个行动的道德价值或许依赖于行动者的动机或蕴含的目标，它们可以归属于这个人的准则。在运用于上述利己主义者时，奥尼尔的提议正确地说明，这个人的行动绝无道德价值。如这个人的准则所揭示的，他从没有由于相信这一行动是其义务而按某种方式行动，也没有出于任何其他道德动机来行动。

然而当我们转向某些其他准则的时候，奥尼尔的提议就失效了。假设在奉行其准则“从不撒谎”时，康德告诉某人真相，并付出了某种他知道自己会付出的巨大代价，因为他正确地相信，他有对这个人告知真相的义务。如果康德是在以这种代价尽自己的义务，且其动机是尽义务，那么这对于给予其行动以

道德价值来说已是绰绰有余。与此无关的是，康德无法意愿其所奉行的准则成为普遍的。只要人们是由于真正相信其行动是自己的义务而尽义务，那就都适用于类似的主张。至于这些人是否在奉行某种他们不能合理地意愿其为普遍的准则，则是无关的。就像一个行动的不当性一样，一个行动的道德价值也不依赖于行动者的准则，这里的准则意指这个人奉行的策略。

我的结论是，我们应当修正康德的公式，以便它们不指涉诸如此类的准则。学完大哲学家们的著作，我们应该力图有所推进。站在巨人的肩膀上，我们或许能比他们看得更远。

第十三章 如果每个人都那么做会怎样？

43 群己困境 (Each-We Dilemmas)

我尽管已主张我们应当修正康德的公式，但仍会继续讨论康德自己的公式。值得说明的是，还有修正这些公式的其他理由，而我的许多主张也适用于修正版。 301

在运用康德的公式时我们会问，我们能否合理地意愿“每个人都奉行某一准则”为真。要回答这个问题，我们必须知道可选项 (alternative) 是什么。如果可选项是除我们之外的每个人都奉行某个坏准则，比如说“付出少于我该付的那一份” (Pay less than my fair share)，我们或许就能够合理地意愿每个人都这么做。另一可选项或许是每个人都继续做他们在做的任何事。而康德的公式由此就错误地允许我们按照许多坏准则行动。如果许多人已经在奉行某个坏准则，那么即使该准则为每个人所奉行，也通常不会产生多大的差别。按照康德公式的最佳版本——这似乎也是康德心中所想，我们应该问的是，我们能否合理地意愿如下情况为真：某一准则，与其无人奉行，不如每个人都奉行。

我们还需要了解，有哪种别的准则是每个人都会奉行的。我们能够合理地意愿每个人都奉行某种坏准则，如果可选项是每个人都奉行别种更坏的准则。因此，我们应该问问是否存在别种较好的准则，这里的“较好”意指我们有更强的理由去意愿“每个人都奉行该准则”为真。 302

在运用于如下三点为真的准则或行动时，康德的公式最行之有效：

众多的人奉行该准则或按此方式行动，这是可能的；

无论按此方式行动的人之数量如何，每一行动的效果相仿（similar）；

这些效果在不同的人们之间作大致平等的分配。

在讨论这些情况的过程中，我会用“我们”（we）来指某一群体中所有人；在也适用于女性的意义上使用“他”（he）和“他本人”（himself）。我们通常是某个群体的成员，对于该群体来说如下情况为真

如果我们每一个人^①而不是无人做某种意义上较有利之事，我们就会是做同样意义上较不利之事（if each rather than none of us does what would be in a certain way better, we would be doing what would be, in this way, worse）^②。

我们可将这样的情况称作群己困境（each-we dilemmas）。

考虑每一个人的行动会有利于一个或更多人的情况就足够了。群己困境的一大类别是自利的困境（self-benefiting dilemmas），这常常令人遗憾地被称作囚徒困境（prisoner's dilemmas）。在这样的例子中，我们是某群体的成员，对该群体来说如下情况为真：

（1）我们每一个人各自能够要么自利，要么给予他人某种更大的利益；

① 本章中everyone与each或each person有较明显的区别：前者强调的是—一个群体中的所有人，着重于“无一例外”意思；后者强调的是“各自”，侧重于个体的含义。因此，下文将everyone与each或each person分别译为“每个人”与“每一个人”。虽仅一字之差，但对于理解原文来说却是不可少的。

② 孤立地看这段话可能令人费解，其含义可能要随下文的解释才能逐渐清晰。鉴于作者力求—种高度概括而准确的表达方式，此处不考虑为照顾可读性而转译（为慎重起见，附上原文）。要之，作者大致要表达这样一种悖论：就群体中的某事而言，它从单个成员的角度看是有利的，而且在大多数人不做而个别人这么做时它对这么做的人也是有利的（如搭便车）；但从整体或长远的效果看，它对该群体所有人都是不利的。这种悖论包括但不止于自利的困境或者说囚徒困境，还包括其他很多社会合作情况，涉及公共善的供给、外部性、累积效应等问题。由于本章类似的表达较多，这里以此为例做个统一的说明，以便阅读。

(2) 这些更大的利益在所有这些人之间会作大致平等的分配；

以及

(3) 每一个人所做之事绝不会对其他人所做之事产生显著的影响。

在这样的情况下，如果我们每一个人都利己，那我们每一个人都在做肯定对自己来说是较有利之事，而无论其他人怎么做。但是，如果我们所有人而不是无人这么做，我们就是在做肯定对我们所有人来说都较不利之事。我们无人会获得较大的利益。这样的情况是群己困境，其意思是

303

如果我们每一个人而不是无人做对自己较有利之事，我们就在做对我们每一个人较不利之事。

换个方式倒过来说，

如果我们做对每一个人较有利之事，每一个人就在做对自己较不利之事。

如果我们大多数人这么做，那对我们也是不利的，并且这么做的人越多越不利。这样的主张并非关于那些具有误导性称呼的重复的囚徒困境（**repeated prisoner's dilemmas**），如我的一个注释所示，那些困境的重要性要低得多。

群己困境尽管常常被忽视，但极为普遍。更确切地说，诸如此类的情况仅涉及两个人或几个人，那是极少出现的；而众多的情况都涉及众多的人。

许多诸如此类的情况可称作贡献者困境（**contributor's dilemmas**）。它们涉及公共善（**public goods**）：即使那些未促进其产生的人也可从中受益的成果（**outcome**）。其中一些例子是清新的空气、国防，以及法律与秩序。在许多这样的例子中，如果每个人都对这样的公共善有所贡献，那对每个人来说都比无人作贡献要有利一些。但对每一个人来说，如果他本人不作贡献将更有利。他将免于付出其个人成本，而从他人那里获取较大利益的可能性丝毫不会降低。在许多这样的情况下，公共善在于我们避免会对每个人都不利的结果，而所需的贡献不是财物方面的，而是某种形式的自制（**self-restraint**）。

有无数这种类型的实例。例如，在渔夫困境（fisherman's dilemmas）中，如果每一渔夫都使用较大的渔网，那么不管其他渔夫怎么做，他都会多捕鱼。但是，如果所有渔夫都使用较大的渔网，鱼的存量就会下降，乃至不久他们就全都少捕鱼。而如下情形仍为真：对每一渔夫而言他若用大网会更有利，以及他们如果都这么做就会全都捕甚至更少的鱼。某些其他情况涉及诸多做法合在一起造成污染、堵车、森林滥伐、过度放牧、水土流失、旱灾、人口过剩。

这些情况之所以常被忽视，是因为在许多此类的例子中，上述主张不适用于其中有些人。比如说，或许有些技术娴熟的渔夫即使在过度捕鱼的情况下，仍会捕到同样多的鱼。但即便如此，其他渔夫仍面临群己困境。在我对这种情况的描述中，“每个人”意指“某群体的所有成员”。断言（1）至（3）可适用于某些人群，即便这些人群中的有些人属于同一个社群（community）且以类似的方式行动，但并不是这个群体的成员。

许多群己困境不涉及在自利或给予他人较大利益之间做选择。只要人们有着不同和部分冲突的目标，上述情况就可能出现。情况可能是这样的：如果我们每一个人而不是无人做最有利于实现其自身目标之事，那就更不利于每个人实现其目标。其中有些或许是道德上要求的目标。根据常识道德（common sense morality，可称为M），我们有特别的义务（obligations）将某些利益给予那些与我们有某种关系的人。诸如我们的孩子、父母、学生、病人、客户、同事、消费者或我们所代表者，都属于这样的相关者。可称之为我们的M相关者（M-related people）。我们如果应当对这些人的福祉给予某种优先性，就可能面临群己困境。比如在父母困境（parent's dilemmas）中，我们每一个人可能要么为自己的孩子谋利，要么给予他人的孩子以较大的利益。如果我们每一个人而不是无人优先考虑为自己孩子谋利，那就对我们所有的孩子都更不利。许多这样的困境引发于自利的困境。例如，可怜的渔夫们全都少捕鱼，此时或许不仅更不利于他们，而且也更不利于他们营养不良的孩子们，这些孩子的营养甚至会更差。

群己困境对实践和理论两方面都提出了难题。对于某些情况，实践难题已得到至少是部分的解决。有些解决方法是政治的，关涉改变我们的处境（situation）。例如，在许多关于公共善的例子中，通过不可避免的或是实施惩罚逃税的征税，不作贡献要么变得不可能，要么对每一个人更不利。然而，在其他许多情况下，政治的解决途径不可能实现，或成本过高。就其中有些情况而言，我们已经达成了心理的（psychological）解决方案，这里的“心理

的”意指：在不改变处境的条件下，我们所有人或大多数人选择给予他人以较大的利益。这样的解决方案通常依赖于我们拥有并奉行某种道德信念。我们尽管自己要付出成本，但仍会促进某些公共善，因为我们相信应当作贡献。

在群己困境的道德的解决方案之中，与此处尤为相关者有二。我们或许是行动后果论者（Act Consequentialists），认为应当总是给予他人以较大的利益，因为我们由此将实现更大的善。我们所有人如果都奉行这种道德信念，那就全都会去促进诸如此类的公共善。但这样的解决方案很少实现，因为既是行动后果论者又经常奉行其道德信念的人为数极少。

还有康德主义的解决方案。如果无人促进诸如此类的公共善，那对于我们所有人来说就比每个人都去促进要不利得多。我们不能合理地意愿如下情况为真：每个人而不是无人奉行“不作贡献”的准则。因此，我们如果都是有良知的康德主义者，总是奉行康德的自然法公式，那就都会去促进诸如此类的公共善。

对于某个贡献者困境，一旦我们已经达成某种道德的解决方案，常识道德就会要求每个人都继续作贡献。在此情况下，通常会有某些搭便车者（free riders）：受益于公共善而又不作任何贡献的人。对于每一搭便车者来说，其本人在一定程度上给他人强加了一个更大的总负担（a greater total burden）。常识道德将这样的做法谴责为不公平。而正是这样的一些例子之中，我们拷问“如果每个人都那么做会怎样？”，可以说是最恰如其分。

在未解决的群己困境中，情况稍有不同。如果无人贡献于某种仅为可能的公共善，那就无人搭便车或是没分担其合理份额。但康德的自然法公式仍蕴含着：每个人都由于没作贡献而行动不当。这个公式原本可以成为针对如此情况的特别设计。如果每个人都没作贡献，那我们的对话就不能是：“如果每个人那么做会怎样？”因为每个人都是那么做的。但我们可以换个方式提问。相较于每个人都作贡献而获得公共善的世界，我们不能合理地意愿“无人为了获得公共善而作贡献”的世界为真。因此，康德的公式要求我们所有人都作贡献。

306

在运用于诸如此类的例子时，康德的公式冲突于某些被广泛持有而又至少有部分错误的道德信念，并且可以引导我们去修正后者。在未解决的群己困境之中，大多数人相信：与其给予他人以大利，不如让我们自己或与我们M相关者得小利；这对我们来说或是被允许的，或是被要求的。按照康德的自然法公式，这么做是不当的。我们无人能够合理地意愿“我们所有人而不是无人继续依此方式行动”为真，因为那就对我们所有人或所有与我们M相关者更不利。

康德的公式不仅冲突于某些被广泛持有的信念，而且以尤为有力的方式

挑战着这些信念。尽管行动后果论者也会主张，每个人都应当给予他人以大利，但我们更难拒绝对这一结论的康德主义论证。在未解决的群己困境中，我们每一个人都将力图有益于我们自己、我们的孩子、父母、学生、病人或与我们M相关者。从个人的层次来判断，我们每一个人都成功，因为我们每一个人都是在做对他自己或他的孩子、父母、学生、病人等更有利之事。但是，我们是在做对所有这些人都更不利之事。即便就我们自己而言，我们也是失败或做更不利之事，因为我们在使如下情况成真：将更不利于实现每个人为道德所要求的目标。在如此情形之下，我们奉行常识道德原则的方式是直接地集体自我挫败的（*directly collectively self-defeating*）。如果我们是合理利己主义者（*Rational Egoists*），那也丝毫不会对上述观点有所反驳，因为这种形式的利己主义是一种关于个人的合理性与理由的理论。而道德原则或理论旨在回答关于我们所有人应当做什么的问题。因此，这样的原则或理论一旦在集体的层次是直接地自我挫败的，就显然失败并陷入自相谴责。

康德接近于给出上述论证。在讨论利他义务的程度时，康德写道：

以自我幸福的牺牲来促进他人的幸福，这一准则……如果变成一条普遍法则，那将自相冲突。

康德的意思必定是“以自我幸福的更大牺牲”。其要必在于，每个人如果都以自我幸福的更大牺牲来促进他人的幸福，那每个人的幸福之所失将会大于其所得。如果如此做法的效果在不同的人之间作大致平等的分配，该要点就是正确的。此即该准则如何会“自相冲突”之所在。类似的看法适用于“以他人幸福的更大牺牲来促进自我幸福”的准则。根据类似的假设，该准则如果成为一条普遍法则，也会自相冲突。仅有一个准则能够不自相冲突地或不集体自我挫败地成为普遍的，此即“做总体上会最大地促进每个人的幸福的任何事”的准则。

在运用于一类未解决的群己困境时，康德的公式具有甚至更大的价值。在许多情况下，

- 307 （4）我们每一个人可能以使自己或与我们M相关者受益的方式，给他人带来更大的总负担。但这些负担将为极多的人所分摊。因此，每一次这么做强加给其他人的负担，分到每一个人身上将微不足道，通常难以察觉。

在大多数此前提及的贡献者困境中，上述主张都是正确的。当我们得知，我们的行动给其他许多人带来了负担，但分给其中每个人的负担是如此微不足道而难以察觉；这并不会唤起我们通常对他人的关切。我们即使是有良知的行动后果论者，也可能忽视这种效果。但是，一旦我们之中许多人都这么做，其综合效果或许是极大极坏的。由于使用化石燃料，我们鲁莽而自私地使地球大气过暖，这种方式即为其中一例。在诸如此类的例子中，康德的自然法公式可以起道德放大镜之用，使我们得以看清我们在做什么。我们不能合理地意愿如下情况为真：我们一起给自己、我们的孩子，以及我们孩子的孩子施加这样的伤害。

44 门槛反驳 (The Threshold Objection)

现在我们可以转向某些康德公式于其中不那么有效的情况。根据康德的

自然法公式：按照某准则行动是不当的，除非我们能够合理地意愿“每个人都奉行该准则”为真。

但在某些情况下，某一行动是否不当，取决于有多少人按此方式行动。在此点为真时，康德的公式就会由于谴责正当的行动或允许不当的行动而失效。

在对诸如此类情况的讨论中，考虑如下行动就足够了：该行动的正当性，部分地取决于可预期的效果。对于许多准则来说，如下情况为真：

(5) 如果过多的人按此准则行动，这些人的行动就会产生不良效果；但一旦较少的人按此准则行动，其效果就是中性的或良好的。

由此如下情况可能为真：

(6) 如果过多的人按此准则行动，这样的行动就会是不当的；尽管如此，一旦较少的人按此准则行动，这样的行动就是可允许的，甚至可能是为道德所要求的。

在这样的情况下

(7) 我们大多数人不能合理地意愿“每个人都奉行这样的准则”为真。

当这样的行动是可允许的甚或为道德所要求的时，康德的公式会错误地予以谴责。

309 一个例子是“不生养孩子，以便毕生致力于哲学”的准则。康德如果奉行此准则，那并非行动不当。但他不能合理地意愿“每个人都奉行该准则”为真。因此，康德的公式似乎蕴含着，康德故意不要孩子本该是不当的。再想想如下准则：“消费而不生产食物”、“成为一名牙医”，以及“在冰岛（Iceland）生活，以汲取萨加传奇（Nordic Sagas）^①的精髓”。在现实世界中，奉行这些准则并非不当。但既然我们不能合理地意愿“每个人都奉行这样的准则”为真，那康德的公式就似乎蕴含着，诸如此类的做法是不当的。其他例子有：“不吃第一片蛋糕”、“等他人说完再发言”，以及“与另一辆车狭路遭遇时，停车等待直到那辆车通过”。我们不能合理地意愿“每个人都奉行这些准则”为真。在这样的世界，大多数蛋糕永远没人吃，大多数交谈永远不会开始，而许多人的旅程永远达不到目的地。但在现实世界，奉行这些准则并非不当。

既然该问题是由这样的行动而引起的：仅当如此行动的次数超过某个大致的门槛时该行动才是不当的，那我们可以称之为门槛反驳（Threshold Objection）。

博格（Pogge）提出，要答复对康德观点的这一反驳，我们应该从康德的自然法公式转向其道德信念公式。我们尽管不能合理地意愿“每个人按此类准则行动”为真，但能够合理地意愿“每个人都相信这样的行动是道德上可允许的”为真。即使每个人都持有这些信念，也不会产生过多的人选择按这些方式行动的危险。大多数人已相信，他们按照我刚才提到的准则行动是可允许的。但仍有足够多的人生养孩子、生产食物。没有过多的牙医或冰岛居民，也没有过多总是让他人先吃、先说、先走的绅士。既然我们能够合理地意愿“每个人都相信这样的行动是可允许的”为真，康德的道德信念公式就允许这些行动。

① 萨加传奇（Nordic Sagas）指的是古代挪威或冰岛讲述冒险经历和英雄业绩的长篇故事。

我认为，上述主张不是对这个反驳的充分答复。我们如果无人生养孩子，那就会终结人类历史。我们如果无人生产食物，那就会由于让自己和孩子饿死而更残忍地终结历史。这些后果不仅仅是我们不能合理地意愿的。如果全都按这些方式行动，我们就是在不当地行动。我们也不能合理地意愿“每个人都错误地相信如此行动并非不当”为真。这样说是不够的：即使所有人都持有此类错误的信念，也不会产生有过多的人这么行动的危险。我们总是有理由来要求自己与他人不持有错误的道德信念，而且对此不存在任何相反的理由。

博格提出对该反驳的又一答复。仅当我们的行动会产生某种效果的时候，我们才会意图按某种方式行动；在此意义上，许多准则都是有条件的（conditional）。在我们的行动不会产生意图之中的效果或会产生某种其他的负效应时，此类准则就不适用。从这些方面来看，我们的准则或许是隐含地（implicitly）有条件的，即使我们尚未对这些条件有自觉的考量。这么讲就行了：如果未满足这些条件，我们就不会按照这些准则行动，而不会改变我们本来的想法。

在似乎会遭到康德自然法公式错误谴责的实际准则之中，大多数至少是隐含地有条件的。我们如果有意不生产任何事物，那么在挨饿的情况下该意图不会实施。我们的准则类似于“只要在生产食物的其他人足够多，就不生产食物”。我们能够合理地意愿“每个人按此准则行动”为真。因此，康德的公式并没有这样的蕴含：我们没有去生产食物是在做不当之事。

还可以假设，在接受“成为一名牙医”的准则的那些人之中，大多数人仅当能够由此谋生的时候才有意奉行该准则。或许我们能够合理地意愿“每个人都接受这个有条件的准则”为真，因为我们知道，就大多数人而言，该准则的条件不会得到满足。但在此，康德的公式会使我们的道德推理呈现为一种相当古怪的形式。而且我们有某种理由不愿意“每个人都接受这一准则”为真。那会是意愿一个这样的世界：其中所有人都想成为牙医，乃至大多数人会由于抱负不能实现而感到失望，因为没有留给他们的牙医职位。在此更有道理的说法是遵循博格的第一个提议，转向康德的道德信念公式。我们可以宣称，任何人都被允许按照这种有条件的准则行动，因为每个人都能够合理地意愿“每个人都相信如此行动是可允许的”为真。这种方式更佳，可以解释为什么在一个需要补牙的世界之中成为一名牙医并非不当。

我们依然没有完全答复门槛反驳。尽管大多数人的准则呈现为此类有条件的形式，但仍有某些例外。康德或许本来认为，既然可以指望其他大多数人有

孩子，那么他没有孩子就是可允许的。但在那些选择不要孩子的人之中，有些人是按照无条件的准则行动的。而道德原则应当有效地适用于那些仅为虚构的情形，只要这些情形所包含的内容足够清晰。我们可以想象某些极端的、无条件的准则，它们一旦被普遍接受，将导致我们所有人都变成没有孩子的、就业不充分的、挨饿致死的冰岛牙医。既然我们不能合理地意愿“每个人都奉行这些无条件的准则，或都相信如此行动是可允许的”为真，康德的公式就错误地谴责我们对这些准则的奉行，即使是这样的情况——我们知道，很少有人按照这些准则行动，因而我们的做法会取得良好效果——之下也是如此。

然而，这并非一个新反驳。像利己主义的准则“做最有利于己的任何事”和康德的准则“从不撒谎”一样，上述准则都是混合型准则，对它们的奉行有时但并不总是不当的。我已经宣称，要答复这一反驳，我们应该不将康德的公式运用于可能意指策略的准则，而是使之运用于对人们在做之事的道德相关描述。根据我们对康德自然法公式的修正版，

LN3：我们在不当地行动，除非我们在做之事是，我们能够合理地意愿每个人在类似情形下都会尽可能做的。

312 设想在对这些无条件准则的奉行之中，我们将会没有孩子或不生产食物，所处的情形是我们知道并没有过多的人按此方式行动。我们就能够合理地意愿“每个人在类似情形下尽可能地按此方式行动”为真。在这样一个世界，不会有过多的人这么做。因此，LN3就不会错误地蕴含这些行动是不当的。

45 理想世界反驳 (The ideal World Objections)

还有另外一种情况，其中某一行动的不当性可能取决于按这种方式行动的人数。如下情况可能为真

(8) 如果足够多的人按某一方式行动，这些人的行动就会产生良好的效果；但一旦较少的人按此方式行动，其效果就是或可能是很坏的。

由此如下情况可能为真

(9) 如果足够多的人按此方式行动而在其他情况下这么做是不当的，那么我们就应当这么做。

许多作家宣称，康德的自然法公式要求某些诸如此类的行动，甚至在它们显然不当时也是如此。

先考察“绝不使用暴力”的准则。不时有人声称，康德的公式要求我们按此准则行动，因为不存在任何与之相冲突的、我们能够合理地意愿每个人都奉行的其他准则。如果这一点为真，康德的公式就会要求我们绝不使用暴力。

和平主义 (Pacifism) 直觉上相当有吸引力，而且很多人 (我父亲是其中之一) 是基于康德主义理由的和平主义者。但像必须绝不撒谎的康德信念一样，和平主义过于简单。回到二战 (the Second World War) 期间来看。如果除德国人之外每个人都是和平主义者，那将允许希特勒统治世界，其结果甚至可能比这场恐怖的战争更糟。如果康德的自然法公式蕴含着反抗希特勒的军队是不当的，那将对这个公式不利。

接着设想

313

失误情形 (Mistake)，数个人的生命处在危险之中。你和我必须在两种做法之间做出选择。其可能结果如下：

	我做A	我做B
你做A	我们救了所有人	我们没救到任何人
你做B	我们没救到任何人	我们救到一些人

我们应当都做A，因为这是救所有人的唯一途径。但假设你由于误解了我们的处境，做了B。尽管知道你已犯下这一失误，我仍然做A，其结果是我们没救到任何人。我知道，我做A将阻止我们去救一些人，这些人是若我做B则我们本来可以救的。但作为一名康德主义者，我认为应当做A，因为这是我能够合理地意愿我们两人都做的唯一之事。

如果康德的公式蕴含着：尽管知道你已经做B，我仍应当做A，那么这一蕴含是完全不可接受的。虽然和平主义有些道理，但这样的主张是荒谬的：我在此应当做A，由此任凭某些我们本可以救的人死去。

这些例子表明了对康德的自然法公式的又一种反驳。科丝嘉写道，康德的“行为标准”（standard of conduct）

是为一种理想事态（state of affairs）设计的：我们应当总是像仿佛生活在目的王国（Kingdom of Ends）那样行动，无论可能出现怎样的灾难性结果。

314 科丝嘉认为，这种问题将由有人行动不当这一事实而引发。但如失误情形所表明的，对康德公式的这一反驳不是仅由故意做错而引发。尽管这个例子设计得很简单，但存在许多类似的实例。如下情况经常是真的：如果像有人宣称的、康德的公式所要求的那样，我们去做我们能够合理地意愿每个人都应做的事，我们的行动就会可预期地产生某种负面效果，这种效果使行动变成不当的。在讨论此类情况时，希尔写道：

问题在于，在这个替他人设定规则的世界之中行动，其结果可能证明是灾难性的。

根据我们可称为的

理想世界反驳（Ideal World Objection）：康德的公式错误地要求我们按照某些方式行动，即使是在这样的情形之下：其他某些人没有按这样的方式行动，因而我们的行动将让事情变得非常糟糕且并非出于任何良好的理由。

对于这一反驳的讨论，考虑如失误情形之中的情况——最佳的状况是所有相关者都以同样的方式行动——就足够了。考察如下准则（简称M）：

M1：做我能够合理地意愿每个人都会做的任何事。

根据理想世界反驳，与意愿每个人都按M1行动相比，我们不能合理地意愿无人按M1行动。如果这一断言是正确的，那么康德的公式就要求，即使如失误情形那样，即使行动可预期的效果非常糟糕，我们也要按照M1行动。

然而，上述断言不正确。一个改进准则如下：

M2：做我能够合理地意愿每个人都会做的任何事，除非其他某个人没有这么做：在此情形之下，做我能够合理地意愿其他人在此处境中会做的任何事。

我能够合理地意愿“每个人都按照M2行动”为真。在失误情形中，我们俩如果都做A，就会都按M2行动，因为这是能救所有人生命的方式。但我知道你没有依此行动，因为你已错误地做了B。鉴于你的失误，我就不能合理地意愿我做A，那样会让我们救不了任何人。要遵循M2，我就必须做B，由此使我们能够救至少一些人。康德的公式既然允许我奉行M2而不是M1，那就允许我以显然正当的方式对你的失误做出回应。

下面回到和平主义的准则“绝不使用暴力”。根据理想世界反驳，康德的公式要求我们按此准则行动，因为不存在任何与之相冲突的、我们能够合理地意愿每个人都会奉行的其他准则。如前所示，这一断言不正确。一个改进准则如下：

315

绝不使用暴力，除非其他某些人已使用侵略性的暴力——在此情形之下使用有节制的暴力，只要它是我自卫和保护他人的唯一可能方式。

每个人都能够合理地意愿“每个人都奉行该准则”为真，因为这将产生一个永远无人使用暴力的世界。因此，康德的公式不要求我们成为和平主义者，而是允许我们使用有节制的暴力以抵制侵略。

类似的断言适用于所有此类的情况。康德的公式从未要求任何人奉行像M1那样的无条件准则，或和平主义准则。每个人都能够合理地意愿“每个人都奉行像M2那样的有条件准则或抵制侵略的准则”为真。如康德的公式所允许的，在奉行如此准则的过程中，我们能够以最佳的方式对他人的不当做法或失误做出回应。

然而，还存在另一问题。康德的自然法公式只是允许我们奉行这些改进准则。考察如下准则：

绝不使用暴力，除非其他某些人已使用侵略性的暴力——在此情形之下，我要杀死尽可能多的人。

如前所示，每个人都能够合理地意愿“每个人都奉行该准则”为真，因为这将产生一个永远无人使用暴力的世界。但在现实世界，某些人已使用侵略性的暴力。既然该准则通过了康德的检验，康德的公式就允许我们其余的人以杀死尽可能多的人的方式来奉行之。接着考察：

守诺并帮助需要帮助者，除非其他某些人没这么做——在此情形之下，效仿他们的做法。

316

这个准则也通过了康德的检验。每个人都能够合理地意愿“每个人都奉行该准则”为真，因为这将产生一个每个人都守诺并帮助需要帮助者的世界。而在现实世界，有些人没有依此行动。既然该准则通过了康德的检验，康德的公式就错误地允许我们余下的人，通过违背一切许诺和绝不帮助需要帮助者来效仿那些人。

为了以更简明的方式来说明这个问题，我们可以求助于

M3：做每个人都能够合理地意愿每个人都会做的事，除非其他某些人没这么做——在此情形之下，做我想要做的任何事。

由于每个人都能够合理地意愿“每个人都奉行M3”为真，M3就通过了康德的检验。我们知道，现实世界之中有些人没有奉行M3，因为这些人没有做每个人都能够意愿他们会做的事。因此，通过允许我们奉行M3，康德的公式就允许我们其余的人做想要做的任何事。

根据理想世界反驳，康德的公式有时要求我们要像仿佛处于一个理想世界那样行动，即便是在这样的情况下——现实世界之中这样的行动会产生灾难性的效果且显然是不当的——也是如此。我们经常出于其他理由而必须将康德的公式运用于有条件的准则，若如此，则我们可以答复上述反驳。但我们现在发现，在运用于此类准则时，康德的公式所提的要求太少。根据这个：

新的理想世界反驳 (New Ideal World Objection)：一旦少数人没有做我们能够合理地意愿每个人都会做的事，康德的公式就不再蕴含着任何行动是不当的。

如果不能答复这一反驳，那将产生至少具有同等破坏性的结果。

类似的主张适用于其他某些道德原则或理论。根据规则后果论（Rule Consequentialism）的一个版本，或者说：

RC：每个人都应当遵循这样的规则：每个人都遵循该规则，将使事情进展最好。

如果做到了某一规则要求我们做的事情，我们就遵循了该规则。经常有人提出反驳说，RC要求我们遵循这些理想规则（ideal rules），即便我们知道其他某些人并没有遵循这些规则，因而我们的行动将产生灾难性的效果。可以答复这一反驳。考察：

317

R1：遵循这样的规则：每个人都遵循该规则将使事情最顺利地进行；除非其他某些人没有遵循该规则——在此情形之下，鉴于他人的行动，去做将使事情进展最好的任何事。

这是理想规则之一，因为每个人都遵循R1将使事情进展最好。由此RC就没有要求我们去遵循这样的理想规则：若该规则的遵循者仅为一部分人，则将产生灾难性效果。但考察：

R2：遵循这样的规则：每个人都遵循该规则将使事情进展最好；除非其他某些人没有遵循该规则：在此情形之下，做你想要做的任何事。

R2也是理想规则之一，因而RC允许我们遵循R2。我们知道，现实世界之中有些人没有遵循这样的理想规则。因此，通过允许我们遵循R2，RC就允许我们其余的人做想要做的任何事。类似的反驳适用于规则后果论的大多数其他版本，譬如这样的理论：它们诉诸其为每个人或大多数人所接受将使事情进展最好的规则。而且类似的反驳还适用于某些契约论的（Contractualist）道德理论。

要答复对康德自然法公式的这一新的反驳，我们就应该再次修正这个公式。在将这个公式运用于某一准则时，这样提问还不充分：我们能否合理地意愿“每个人都按此行动”为真。康德的公式可以成为：

LN4: 按照某一准则行动是不当的, 除非我们能够合理地意愿该准则为每个人或任何其他人数的人们所奉行, 而不是无人奉行。

318 某一准则要通过这一更宽松的 (wider) 检验, 我们就必须能够合理地意愿该准则的奉行者——不仅是每个人而不是无人 (by everyone rather than no one), 而且是大多数人而不是无人、众多人而不是无人、若干人而不是无人, 以及任何其他人数的人而不是无人。我们必须能够合理地意愿: 无论有多少人没有按此准则行动, 其他每个人都按此准则行动。

康德的公式如果按此方式被拓宽, 就会谴责我们已讨论过的坏准则。一个例子是:

不使用暴力, 除非其他某些人已经使用侵略性的暴力: 在此情形之下, 我要杀死尽可能多的人。

我们尽管能够合理地意愿“每个人都奉行该准则”为真, 但不能合理地意愿任何其他人数的人奉行之。如果有任何人使用侵略性的暴力, 其他每个人就会按此准则行动, 杀死尽可能多的人。

当我们思考许多准则和行动的时候, 康德公式的这一版本不会产生任何差别。有许多行动总是正当的, 无论按此方式行动的人数如何。在这种情况下, 存在着我们能够合理地意愿任何人数的人都应奉行的、无条件的准则。一些例子是准则“帮助需要帮助者”和“绝不仅仅为一己之便而去伤害他人”。然而如我们所见, 在考虑某些其他种类的行动时, 我们所能够合理意愿的, 就是人们奉行有条件的准则, 这些准则告诉我们要虑及他人的行动。某些此类的准则可能采取这种形式:

做A; 除非做A者的人数或比例低于或将会低于某个门槛, 在此情况下做B; 或者低于其他某个门槛, 在此情况下做C。

这些门槛之中, 有些可定义为做A者的人数或比例; 若低于这一人数或比例, A种行动将不再产生某种好效果, 或开始产生某种坏效果。

319 类似的主张适用于规则后果论。此前陈述的公式可以变成

RC2: 每个人都应当遵循这样的规则: 任何人数的人而不是无人遵循这些规则, 将使事情进展最好。

某些规则可能采取此类有条件的形式。这些规则会告诉我们, 我们应当鉴于正在遵循这些规则者的人数或比例, 按照会使事情进展最好的方式行动。类似的主张将适用于那些诉诸“如果人们接受某些规则那会怎样”的RC版本。

这个版本使得规则后果论在某些方面更接近于行动后果论(Act Consequentialism)。在问如下问题时其正确性尤显重要: 世界上的富人应当将其收入或财富的多大份额, 给予现在每天靠约2美元过活的10多亿人? 在应用于该问题时, 规则后果论的大多数版本没有提出多高的要求。这些理论诉诸的相关主张是, 若所有人或大多数人接受或遵循某些原则, 则什么是正确的。如果所有或大多数富人给予穷人以其财富或收入的某个相当平常的份额, 比如1/5乃至1/10, 这或许是最佳事态。这会有很大的改观, 因为如今最富裕国家所给予的还不到1/100。如果我们将“所有”或“大多数”变为“任何人数的人”, 以此方式来修正规则后果论, 从而诉诸刚才提及的有条件的规则, 那么规则后果论通常会提出高得多的要求。如果大多数富人所拿出的并没有达到富人应当给出的最佳份额, 那么最佳规则就会要求其他富人拿出一大笔。

在按此方式修正康德的*自然法公式*的过程中, 我们放弃了“如果每个人都那么做会怎样”这一问题所表达的观念。但是, 该观念成功使用的可能性, 仅限于某些种类的情况。在群己困境之中, 我们如果是没有促进公共善的搭便车者, 那就可能受到“如果每个人都那么做会怎样”这一问题提出的正当挑战。但在众多其他的情况下, 回答“大多数人不会”就足够了。

康德的道德信念公式所诉诸的是某种不同的观念, 它或许可以成功地用于一切情况。我们尽管不能言之有理地认为, 每个人都应当奉行同样的准则或按同样的方式行动, 但能够言之有理地认为, 每个人都应当持有同样的道德信念。因此, 人们在反对我们的一个道德信念时说“如果每个人都像你那么想会怎样”, 此时我们简单地答以“大多数人不会”是不够的。如果我们不能合理地意愿“每个人都相信某种做法是可允许的”为真, 那么这一事实可能就如康德所认为的, 表明这样的做法是不当的。

现在我们可以转向某些较为简单而又更为基本的问题。

320

第十四章 不偏不倚

46 黄金规则

321 在叙述其自然法公式如何解释我们的利他义务（duty to benefit others）时，康德写道

我希望其他每个人都有益于我；因此，我也应当有益于其他每个人。

这或许让我们想起

黄金规则（The Golden Rule）：己之待人，当如希望人之待己。

这条规则所表达的，或许是最广为接受的基本道德观念，它分别被各自发现于至少三种最早的世界文明。康德尽管将其公式称作“道德的最高原则”，但将黄金规则嗤之为“微不足道”，且不适宜于作为普遍法则。这一规则应该遭到康德的轻视吗？

对于黄金规则的拒绝，康德写道：

它不可能作为一条普遍法则，因为它不包含对自己的义务之根据，也不包含爱他人的义务之根据（因为许多人会愉快地同意，他人无需有益于己，只要自己可能免于有益于他人）；而最后，它不包含对他人应尽的义务之根据，因为一个罪犯会以此为据，来抗辩要惩罚他的法官。

根据康德的一个反驳，黄金规则不蕴含我们的利他义务。康德声称，许多人会愉快地同意，绝不受惠于他人。

这一反驳适得其反。黄金规则意味着，如果这些人自己想要得到帮助，那么他们就应当有益于或帮助他人。康德并不否认这些人想得到帮助。他做出了不同的断言，这些人会同意不要帮助，只要他们由此将免于帮助他人。用康德的语言来陈述，这一断言就是这些人将意愿“不帮助他人的准则成为普遍法则”为真。这并不意味着，根据黄金规则，这些人没有帮助他人的义务。是康德的公式而不是黄金规则，允许我们奉行我们能够意愿其为普遍法则的准则。

我们可以修正康德的反驳。他或许要求我们去考虑那些不想得到他人帮助的人，他们是否会由此免于帮助他人。那么康德可能宣称，既然这些人不想得到帮助，黄金规则就没有蕴含他们有帮助他人的义务。

然而，如前所示，这一反驳将适用于康德自己的公式。根据该公式，这些人如果不能意愿“不帮助他人的准则成为普遍法则”为真，就应当帮助他人。这些人如果甚至不想得到帮助，就能够更轻易地意愿该准则成为普遍法则。康德宣称，无人能够意愿这样的法则，因为这样的人会由此“剥夺自己得到其想要的帮助的一切希望”。这一断言不适用于那些不想得到帮助的人。

康德可能答道，就不希望得到帮助而言，这些人是不理性的。而他可以接着论证说，在运用于这类人的时候，其公式之表现优于黄金规则。康德可能宣称，黄金规则既然诉诸这些人的不理性的欲望，就没能蕴含他们有帮助他人的义务。相比之下，这些人不能合理地意愿“他们将绝不接受帮助”为真。因此，康德的公式的确蕴含着他们有这种义务。

对于黄金规则的这一反驳是无力的。我们可以先解释，为什么在其所述的大多数版本之中，这一规则没有诉诸我们会怎样去意愿他人待我。我们并非绝对的君主或独裁者，能够随心所欲地决定他人的行动方式。既然对他人没有这样的权力，我们只能想要或希望他人确实按某种方式行动。康德的公式要求我们去想象或假设，我们拥有选择或意愿“他人确实按某种方式行动”的权力。黄金规则可以采取同样的形式。它无需诉诸我们的欲望，而是能够诉诸我们会怎样去意愿自己被对待——或我们会怎样在意愿（*would be willing*）被对待，如果我们有选择的话。黄金规则的某些耳熟能详的陈述如“己之所欲，施之于人”（*Do as you would be done by*），已经采取了这种形式。

黄金规则还可以诉诸我们会合理地（*rationally*）选择或意愿什么。确实如

通常所言，这个规则没有使用合理的（rational）这个概念。但在康德对其公式的诸多陈述之中，仅有两个陈述使用了这个概念，而且没有一个明确地诉诸我们能够合理地意愿什么。鉴于康德的某些其他主张，他显然想要我们问，我们能够合理地意愿或选择什么。黄金规则（G）可以采取同样的形式，它可陈述为

G2：应仅以如此方式待人：我们会合理地意愿他人用同样的方式来对待我们。

在运用黄金规则的时候，有时这样问就够了：在现实世界，我们是否意愿被待以某种方式。比如说，施虐者不会意愿将要受虐。但在考量诸多种做法时，我们必须问的是，在某些仅为想象的情况下，我们将要如何意愿被对待。例如，在可以给某个挨饿者以食物时，这样问是不够的：我们是否意愿将不被给予任何食物？如果我们刚吃完美餐且厨房里食物储备充裕，我们对这个问题的回答或许就是“是的”。我们应当问的是，我们即使在挨饿，是否也会意愿将不被给予任何食物。

接下来考察这样一个白人种族主义者：在美国南部种族歧视最严重的时期，他的旅馆拒绝接待黑人。这个人可能宣称在遵守黄金规则。他会说：

应当仅以如何意愿人之待己的方式待人。我允许任何非黑人进我的旅馆。我意愿将被待以这种方式。我实际上是被待以这种方式。我既然不是黑人，就可准入所有旅馆。

324

这段话是对黄金规则的误解。根据该规则，这个人对待黑人的方式仅应当如此：他如果处于他们的境地（in their position）会意愿如何被对待。他必须想象，要么（1）属于黑人的所有旅馆都拒绝接待白人，要么（2）他本人是个黑人。（1）只是其处境的变化，（2）则是他本人的改变。在将黄金规则运用于许多其他情况时，想象中的变化必须发生在我们自己身上，因为对于会受我们行动之影响的那些人，我们必须通过拥有他们的欲望、态度，以及其他生理的或心理的特征，来想象自己在相关方面与他们类似。例如，某个男人要想象得到如他对待女人一般的对待，那他或许就必须想象他是个女人。类似的主张适用于施虐-受虐狂。

那么，黄金规则的较完整陈述就可能是

G3: 应仅以如此方式待人：我们如果处于他人的境地且在相关方面与他们类似，将合理地意愿受到与此一样的对待。

“将会意愿”（**would be willing**）这一说法可能起误导作用。在运用**G3**时，我们不应该这样问：如果处于他人的境地，那时我们会意愿受到怎样的对待；应该问的是：如果随后将处于这些人的境地，我们此刻会意愿随后受到怎样的对待。（如果我类似地问道，“你愿意让你的器官在你死后为他人所用吗”，那我的提问并不是要预料你死后的欲望，而是要你现在的决定。）

康德给出了对黄金规则的另一反驳。康德声称，通过诉诸这一规则，“罪犯可能抗辩要惩罚他的法官”。在此，康德必定是假定这个罪犯会这么说：“你既然不会想要受惩罚，那就不应当惩罚我”。这一反驳是将黄金规则看做

G4: 应如此对待每一个他人：我们如果处于这个人的境地且在相关方面与他类似，将会合理地意愿受到同样的对待。

325

康德对这个规则的拒绝将是正当的。假设在

情形1：我可能救一命，要么是布鲁（Blue）的，要么是布朗（Brown）的。

通过诉诸**G4**，布鲁可以争辩说，我应当救她。我要是处于布鲁的境地，就不会意愿被弃之不顾而死去。布朗可以同样地争辩说，我应当救她。因此，**G4**错误地蕴含着，由于不能如我应当做的那样去对待布鲁或是布朗，我无论怎么做都会行动不当。接着假设在

情形2：我有一小条面包，得满足两个挨饿的人。

通过诉诸**G4**，这两个人各自都会争辩说，我应当给她整条面包。

耶稣（Jesus）在诉诸黄金规则的时候是在诉诸**G4**吗？我让这些分享面包的人，耶稣有意暗示这是不当的吗？答案显然是否定的。不应将黄金规则的意思解作**G4**，而应解作

G5: 应如此对待其他的人们：我们如果处于所有这些人的境地且在相关方面与他们类似，将会合理地意愿受到同样的对待。

然而，该规则的这一改进形式较难应用。我们如何去想象将处于两个或更多人的境地？

对此有若干提议。设想在

情形3: 我可以要么救格林（Green）的命，要么让格蕾（Grey）免于致盲。

326 按照内格尔（Nagel）的提议，我应该想象，自己像变形虫一样随后将分裂成两个人，一个处于格林的境地，另一个处于格蕾的境地。按照黑尔（Hare）的提议，我应该想象，我随后将去体验恰如格林和格蕾一般的生活，并非同时体验两人的，而是一人的紧接另一人的。按照海萨尼（Harsanyi）^①的提议，我应该想象，我会概率均等地处于格林的或是格蕾的境地。按照罗尔斯的提议，我应该想象，我将随机地处于这些人的境地之一。

在将黄金规则运用于某些问题时，我们采取上述提议中的哪一种，这可能会产生差别。但在大多数情况下，这些提议有着同样的蕴含。比如在情形3之中，无论以上述哪种方式来想象我将处于格林和格蕾这两者的境地，我都不会这样去意愿：宁愿在其中一个境地之中免于致盲，也不愿在另一境地之中免于死亡。

在已诉诸黄金规则的那些人之中，许多人或许没有思考G4与G5之间的区别。但如果这些人对这些断言做了比较并明白它们的蕴含，那么他们就会认为，G5更好地表达了其心目中的道德观念。

现在回到康德的主张：通过诉诸黄金规则，一名罪犯可能抗辩说，其法官不应当惩罚他。根据如G5所表达的、对黄金规则的更好解读，法官可以拒绝这一抗辩。法官们应该这样问：他们如果将不仅处于罪犯的境地，而且处于其判

^① 约翰·海萨尼（John C. Harsanyi, 1920—2000），匈牙利裔美国经济学家。理性预期学派的重要代表人物，将博弈论发展为经济分析工具的先驱之一，1994年诺贝尔经济学奖获得者。

决会影响到的所有其他人的境地，那将会怎样去合理地意愿被对待。其他这些人包括如下罪行的可能受害者，即，如果这个罪犯未受惩罚就更可能会发生的罪行，无论是因为该罪犯会被释放而有能力犯其他罪，还是因为他以及其他潜在的罪犯受到威慑的可能性降低。既然这是法官们应当如何运用黄金规则的方式，该规则就没有“无人应该受到惩罚”的错误蕴含。

按照前面引用的那一段中康德余下的反驳，黄金规则不可能成为一条普遍法则，因为它没有涵盖对己的义务。我们可以回答说，该规则既然仅运用于我们如何对待他人，也就没有宣称涵盖对己的义务。然而，如康德在别处提出的，黄金规则的这一特征可能使它错误地描述某些对他人的义务。设想在

情形4：我可以要么救自己的生命，要么使格蕾免于致盲。

如果黄金规则告诉我的只是应当如何对待他人，该规则就可能错误地蕴含着：我应当以我的生命为代价去使格蕾免于致盲。这或许是如果仅处于格蕾的境地，我就会意愿做的事。

要应对这一反驳，该规则可以成为

G6：我们应当这样对待每个人：我们如果将处于所有这些人的境地且在相关方面与他们类似，将会合理地意愿受到同样的对待。

“每个人”一词在此指的是，我们的行动可能影响到的所有人。在许多情况下，我们包含于这些人之中。根据黄金规则的这一版本，在运用于情形4时，我应当做的是，如果处于不仅格蕾的而且包括我的境地，我将会意愿去做的事。如情形3之中，我不会这样去意愿：宁愿在其中一个境地之中免于致盲，也不愿在另一境地之中免于死亡。这个版本更好地表达了黄金规则的假定，即，每个人都同等重要。并不令人惊讶的是，在对该规则的大多数陈述之中，我们仅被告知，要像我们意愿自己如何被对待的那样去对待他人。几乎不存在我们会忽视自身福祉的危险。但对他人的这种指涉（reference）在一定意义上令人误解，因为我们应当以该规则所要求的不偏不倚方式来考虑人们的福祉，而我们包含于这些人之中。

我已论证，康德对黄金规则的轻视不能证明为正当的。但或许仍如康德本人所认为的，其普遍法则公式是一种更好的原则。果真如此吗？

这两种原则通常有着同样的蕴含。而作为道德最高原则的候选，它们俩都满足最明显的要求。在康德的不可能性公式遭到惊人的失败之处，这两种原则大多数情况下都取得了成功。我们大多数人不能合理地意愿“每个人都奉行自利的杀人、伤害、强制、撒谎和偷窃的准则”为真。如果处于受影响者的境地，我们也不会意愿受到这样的对待。

康德的普遍法则公式有两方面类似于黄金规则。就它们的最佳形式而言，这两个原则都诉诸这样的主张，即，关于人们选择什么会是合理的。而且它们都认为，每个人都同等重要，有着平等的道德要求权。奥尼尔写道，康德的公式背后的“直觉观念”是“我们不应将自己挑出来以作特殊的考虑或对待”。

这两个原则的主要差别在于，它们促进我们更加不偏不倚地进行道德思考的方式不同。通过就某些想象中的情况提问，这两个原则都告诉我们去进行某些思想实验（*thought-experiments*）。要运用黄金规则，我们得问“如果那发生在我身上会怎样？”为了运用康德公式的自然法版本和道德信念版本，我们要问的是“如果每个人都那么做会怎样？”以及“如果每个人都相信如此行动是可允许的会怎样？”

在运用黄金规则时，我们的思想实验相当简单。像平常做决定的许多时候一样，我们会问，如果在某一场合我们以某些可能方式之中的每一种去行动，现实世界会发生什么。我们甚至没必要去确定，对于这些特定的可能行动来说，其道德相关的描述是什么。但我们会不仅从自己的、也从可能受我们行动影响的一切他人之视角（*point of view*），去设法思考这些可能性。我们会问，如果处于所有这些人的境地且在相关方面类似于他们，我们会合理地意愿做什么以及对我们做什么。

康德的思想实验在好几个方面都要困难一些。在运用康德的自然法公式时，我们必须先确定我们将要奉行的准则是什么。在我对该公式的修正版中，我们必须确定，对我们行动的道德相关描述是什么。接着我们要比较两种可能世界，或者说世界未来历程的两种可能走向。我们既要问如果每个人都奉行某一准则会发生什么，也要问如果由于每个人都奉行某个别的准则而无人奉行该准则，那会发生什么。类似地，在运用康德的道德信念公式时，我们既要问如果每个人都持有某个道德信念会发生什么，也要问如果由于每个人都持有某个别的道德信念而无人持有该信念，那会发生什么。这四种可能世界或许全都与现实世界极为不同，而且常常难以预测它们会是怎样的。我们可能还得考虑每个人都可能奉行的各种其他的准则，或每个人都可能持有的、可能的道德信

念。然而，从另一种意义上说，康德的公式比黄金规则更易于运用。在问我们能够合理地选择生活在哪种世界时，我们只需从自己的视角来考虑这些世界。

可以将康德的公式和黄金规则与两个别的原则做个有益的比较。根据另一种旧观点，我们应该以某种不同而又较简单的方式来做道德推理。我们应该问，选择或愿意什么会是合理的，其视角既不是从自己，也不是从可能受我们行动影响的他人，而是从某个超然的（detached）观察者，他没有卷入我们正在思考的事例。根据该观念的一个变体，我们会问，想象在其他某个相关方面类似的情况之中每个参与者对我们来说都是陌生人，此时选择或愿意什么对我们来说会是合理的。我们可称之为公正的观察者公式（Impartial Observer Formula）。

我们还可以通过运用康德的同意原则（Consent Principle）来实现不偏不倚。通过提问是否每个人都能够合理地同意某种可能的行动，我们给予每个人拒绝同意的理由以同等的权重。

黄金规则还受到各种各样的反驳。我们处于他人的境地且在相关方面类似于他们，这可能难以想象。而我们必须设法想象的东西经常会是非常不可能的。但如某些作者所宣称的，这并非决定性的反驳。某些思想实验即使要求我们去想象极其不可能的东西，也是有帮助的。爱因斯坦（Einstein）曾不无助益地问道，他如果以光速旅行会看到什么。我们尽管不可能成为正受我们鞭打的马，或被困住的、挨饿的、我们正穿着其毛皮的动物，但还是可以出于道德的用途而对此类事情展开极为充分的想象。

对黄金规则的另一反驳要有力一些。罗尔斯指出，如果想象我们将处于可能受我们行动影响的一切人的境地，那将导致我们忽视如下事实：在现实世界，我们的行动会影响不同的人。某一个人的负担不可能通过另一人的受益来抵消。我们忽略这种“个人的分立性”（separateness of persons），就是在忽略某些事实，这些事实可能给予我们以决定性理由来接受分配正义（distributive justice）原则。

从种种方面看，黄金规则在理论上既不及公正的观察者公式，也比不上康德的同意原则。但就实践的用途而言，这一规则或许是这三种原则中最好的。通过要求我们想象自己处于他人的境地，黄金规则可以提供一种心理上最有效的方式，以使我们变得更不偏不倚且给予我们以道德激励。这一规则已成为世上最广为接受的基本道德观念，其原因或在于此。

我将论证，在这四种使我们变得更不偏不倚的方法之中，康德的普遍法则

公式最不成功。该公式没有谴责众多不当的行动。但我们将会看到，存在对这些问题中的一个康德主义的解决方案。

47 稀罕性反驳与大赌注反驳

人们在行动不当时，或许是在做某种不可能经常做的事。其中一些人能够合理地意愿“每个人都像他们那样行动”为真，因为这样的行动太罕见，不会对他们产生显著影响。我已称之为稀罕性反驳。比如，考察

不公正的惩罚情形 (Unjust Punishment)：除非怀特 (White) 去警局认罪，布莱克 (Black) 就会被定罪且由于怀特所犯的罪行而受惩罚。怀特尽管知道这一实情，但没有任何作为。

331 假定怀特奉行的准则是“让他人顶我的罪而受罚”。要运用康德的自然法公式，我们就得问，怀特能否合理地意愿“每个人都奉行这一准则”为真。要回答这个问题，出于前文^①给出的理由，我们不能诉诸“怀特的行动是不当的”这一信念；也不能诉诸由这个行动之不当性可能产生的道义式理由。如果仅诉诸别的道义式理由，我们或许就得承认，怀特能够合理地意愿“每个人都奉行他的准则”为真。我们可以设想，怀特如果让布莱克顶罪受罚，就会免除多年的牢狱。如果一旦怀特的准则适用，则其他每个人就奉行之，那就会增加这样的风险：怀特随后由于其他某个人的犯罪而受罚。但这一额外的风险很小，而且显然会为怀特免除多年牢狱之利所压倒。康德的公式由此就允许怀特让布莱克顶罪受罚，尽管这个做法显然是不当的。康德的道德信念公式也不会谴责这个做法，因为怀特能够合理地意愿“每个人都相信如此行动为道德所允许”为真。

谋杀性偷窃情形 (Murderous Theft)：在某个沙漠旅行时，格蕾和布鲁都遭蛇咬。布鲁出于慎重带了某种药，它是对这种蛇致命之毒的解毒剂。格蕾救己之命的唯一方法是偷布鲁的药，附带的可预见后

^① 第287页 (原文页码) 有关于道义式信念与道义式理由的论述。后面及附录C还有专门讨论。

果是布鲁死去。

我们可以假设，格蕾知道，无人会揭露他偷布鲁的药，他的生活也不会毁于悔恨。格蕾很年轻，因而可以预期，这么做会给他增加多年值得过的人生。布鲁也可以预期这样的生活，而且还要年轻得多。根据这些假设，所有可行的道德观点都蕴含着：格蕾偷布鲁的药来救自己的命，这是不当的。

先假设，格蕾如果偷药，那就是在奉行“在偷窃是救己命的唯一之法时就偷窃”的准则。格蕾能够合理地意愿“每个人一旦在该准则适用时就奉行之”为真。如下情况不太可能：在如此世界，其他任何人都按此方式来对待格蕾；而且这种风险显然为格蕾救己之命的获利所压倒。根据这些假设，这个例子也支持稀罕性反驳，因为康德的公式会允许格蕾的谋杀性偷窃。

转而假设，在偷窃布鲁的药的过程中，格蕾奉行的是利己主义准则：

332

E：做最有利于己的任何事。

格蕾能够合理地意愿“每个人而不是无人奉行该准则”为真吗？这取决于可选项。如我所言，如果存在某个显然较优的其他准则可供每个人奉行，我们就不能合理地意愿“每个人都奉行某一准则”为真。一个此类的准则可能是：

E2：做最有利于己的任何事，除非这会给他人强加大得多的负担。

如果每个人都总是奉行E而不是E2，那对大多数人来说将远为不利。如我所宣称的，此即利己主义准则通常不能通过康德的检验之原因所在。大多数利己主义者不能合理地选择生活在一个由利己主义者组成的世界。

然而，格蕾却是一个例外。他知道，如果每个人都奉行E而不是E2，他通常就会承受由他人的利己主义行动强加给他的负担。但我们能够合理地设想，即使在这样的世界，格蕾的余生也是值得过的。若如此，则格蕾能够合理地意愿“每个人都奉行E而不是E2”为真。如果每个人都奉行E2，格蕾就不会偷布鲁的药，从而死去。如果忽略道义式理由，我们必定同意，格蕾有充分的理由不去选择他会死于其中的半道德世界，而宁愿选择一个他可以偷药活命的利己主义世界。因此，康德的自然法公式错误地允许格蕾的谋杀性偷窃。出于类似的理由，康德的道德信念公式也是如此。

这些主张支持对康德公式的一种与众不同的反驳。这些公式在此的失败，不是由于他人极少能够奉行格蕾的利己主义准则，而是由于格蕾的不当行动给予自己一个不同寻常的巨大利益。我们可称之为大赌注反驳（High Stakes Objection）。

333 我们或许有几种途径来设法答复这一反驳。例如，我们可以重申罗尔斯的主张：在提问“我们能否合理地选择生活在每个人都奉行某一准则的世界”时，我们应假定该准则已被奉行一段时间，足以让此类行动充分展示其效果。由此就可以论证说，格蕾不能合理地选择每个人总是奉行利己主义准则的世界，因为存在这样一种风险：格蕾已死于这个世界，此前就被其他某个利己主义者杀死。但这个多少有些令人困惑的论证不足以为康德的公式辩护。我们要将该公式与如下三种其他公式进行比较：康德的同意原则、公正的观察者公式，以及黄金规则。而一旦运用于我们现在思考的这类情形，其他三种原则的表现显然好得多。

主要的区别如下。既然布鲁比格蕾要年轻得多，那布鲁之死对她的损失也就大得多。在运用这里的其他原则时，我们会将布鲁大得多的损失纳入考虑。布鲁不会有充足的理由同意格蕾偷药，并由此致自己于死地。任何一个公正的理性观察者如果有选择机会，就会选择格蕾不那么对待布鲁。而格蕾如果将处于不仅是自己的而且是布鲁的境地，那就不能合理地选择被待以这种方式。这三种原则使我们的道德推理更不偏不倚，因此，它们全都正当地谴责格蕾的谋杀性偷窃。

相比之下，在运用康德的公式时，我们忽略了布鲁的福祉，因为我们仅仅是从格蕾的视角来思考。我们的提问是，格蕾能否合理地意愿“救自己的生命且生活在一个利己主义的世界”为真。康德的公式要谴责格蕾的做法，答案必须是否定的。我们必须主张，格蕾不能合理地选择挽救自己生命的那个世界，因为他有决定性的非道义式理由而宁愿选择他死于其中的世界。与我们运用其他三种原则时诉诸的主张相比，这个主张得到辩护的难度要大得多。

48 不可逆反驳

334 对于康德的公式，还存在另一个实践上更重要的类似反驳。黄金规则使我们变得更不偏不倚，其途径是要求我们这样去对待每个人：我们如果处于所有这些人的境地且在相关方面类似于他们，将会合理地意愿受到同样的对待。

康德的自然法公式使我们变得更不偏不倚的途径没有这么直接。在运用该公式时，我们所问的不是“如果那发生在我身上会怎样？”，而是“如果每个人都那么做会怎样？”

这是个有意义的问题。如康德所指出的，我们在不当地行动的时候，通常不公平地使自己成为例外，从而去做我们不想或不愿意他人做的事。康德的自然法公式正当地谴责此类做法。而且我曾讲过，在思考群己困境时，该公式尤为有益。

但康德的问题还不充分。在许多情况下，我们如果不当地行动，就会在自利时给他人带来更大的伤害。黄金规则谴责这样的行动，因为我们不会合理地意愿让他人对我们这么做。但是，在将康德的公式运用于我们对某一准则的奉行时，我们不问，我们能否合理地意愿“其他人对我们做如此之事”为真；而是问我们能否合理地意愿“每个人都对他人做如此之事”为真。而我们或许认识到，即使每个人都对他人做出这些事，但无人会对我们这么做。一旦这一点为真，我们就能够合理地意愿如下情况为真：每个人都像我们一样行动，因为我们由此会从自己的不当行动之中获利，而他人不当的类似行动绝不会给我们强加更大的负担。在最简单的此类情况之下，我们的不当行动是不可逆的，因为我们对他人做的事，他人不可能对我们做。因此，我们可称之为不可逆反驳（Non-Reversibility Objection）。

不同于稀罕性反驳与大赌注反驳，该反驳适用于许多实例。先回顾我们所例举的白人种族主义者。这个人不可能宣称是在遵循黄金规则，但可能宣称在遵循康德的公式。他可以这样说：

在我的旅馆拒绝接待黑人时，我能够合理地意愿每个人都按这种方式行动为真。每个人的确都这么做。每家旅馆的主人都排斥黑人。于是我能够合理地意愿“每个人都相信这样的行动是正当的”为真。如果黑人相信我的行动是正当的，那我就感觉不错了。

335

这个人如果这么讲，那是对康德公式的误解吗？我不是在问，他是否会误解康德的道德理论。康德在某些方面是引人注目的平等主义者，而且康德的观点之中不乏对此类种族主义态度和行动的谴责。我的问题仅在于，康德的自然法公式和道德信念公式蕴含着什么。

在说明其公式时，康德思考的是大多数人不予奉行的准则，以及他认为无

人会希望每个人都奉行的准则。康德想象某个犯错者在问“我能够意愿我的准则成为普遍法则吗”，此时他假定，这个人的准则不是普遍法则。但在某些如这个白人种族主义者的情况之下，该假定失效。这个人的准则已经是一条普遍法则。在奉行“我的旅馆拒绝接待黑人”的准则时，他在做的就是其社会圈子之中所有旅馆主们都做的事。

336 在犯错者奉行此类准则的时候，“如果每个人都那么做会怎样”的提问可能是无益的。如果这些人能够合理地意愿“他们及他人继续那么行动”为真，康德的自然法公式就允许他们的行动。如果对这些犯错者来说，他们及他人现在以某种方式行动是不利的——比如说，在无政府状态或一切人反对一切人的状态下，这可能是真的——那么这些人就不能合理地意愿现存事态或现状持续下去。此时康德的公式会正确地谴责这些人的行动。但在许多情况下，现状对于这些在做不当之事的人来说是有利的。而这种事态之所以可能有利于这些人，部分地是因为他们的坏准则是普遍的，或者是被广泛奉行的。适用于某准则的那些人可能是某种强有力的特权群体，其行动方式是要维持他们相对于其他人的优势。如果他们能够合理地意愿“维持其特权地位”为真，康德的自然法公式就允许这些人的行动。

与以前一样，在设法论证这些人不能合理地选择去维持其特权地位的过程中，我们不应诉诸他们行动本身的不当性，因为由此康德的公式将一无所成。声称可以合理地要求这些人给予其他每个人的福祉以充分的权重，这种主张也不可能有效。当然，康德没有诉诸诸如此类的主张。康德的公式要支持这一观点——即这些人的行动是不当的，我们就必须能够主张：出于其他理由，这些人不能合理地意愿“维持他们相对于其他人的优势”为真。至少就他们之中许多人而论，我们不能言之有理地为这一主张辩护。

求助于康德的道德信念公式也无济于事。正如这些人能够合理地意愿“与他们同一地位的每个人都像他们那样行动”为真，他们也能够合理地意愿“每个人都相信，如此行动在道德上是允许的”为真。这些人没有任何相关的理由倾向于认为，每个人都相信他们的行动是不当的。

比如说考察一下这样的男人：他们将女人看做低人一等、拒绝承认女性的各种权利与待遇、并给女性的福祉以较小的权重，从而有利于己。如果这些男人不能合理地意愿“每个人都像他们一样行动，或者是每个人都相信如此行动是正当的”为真，那么康德的公式就蕴含着如此行动是不当的。这样的主张并不能对这些男人的行动提出适当的反驳。在历史上的大多数时间，包括大多

数女人在内的大多数人都将女人看做低人一等，且相信这样的待遇是正当的。既然不能诉诸这种待遇的不当性，我们就必须承认，许多男人原本能够合理地意愿维持其特权地位。

下面再看奴隶主。康德的公式要谴责奴隶制，我们就必须论证，奴隶主不能合理地意愿如下为真：他们畜养奴隶，以及包括奴隶在内的每个人都相信奴隶制是正当的。由于我们不能诉诸奴隶制的不当性，这些主张可能就难以得到辩护。诉诸康德的同意原则或黄金规则就会好得多。女人和奴隶不能合理地同意被视为低人一等，或仅被视为财产。男人或奴隶主如果处于女人或奴隶的境地，也不会合理地意愿受到这样的对待。

337

类似的主张也适用于强者通过压迫或利用弱者以自利的诸多方式。仅当这些人不能合理地意愿“他们及他人继续以这样的方式谋利，或者是每个人都相信诸如此类的利用是正当的”为真，康德的公式才会谴责他们的行动。由于不能诉诸此类利用本身的不正当性，我们就不能为这些主张提供言之有理的辩护。

作为最后一例，我们可以回到全球不平等问题。在任何一种可行的道德观点看来，控制着绝大多数世界资源的那些人，应当将他们的不少财富或收入移交给世上最穷的人。大多数富人未交分毫。要论证康德的公式谴责这些人的行动，我们就必须主张，这些富人不能合理地意愿“他们及他人继续对穷人不予分毫，或者是每个人都相信富人不交分毫是正当的”为真。由于不能恰当地诉诸这些富人行动的不当性或利他的合理要求，我们就不能言之有理地为这些主张辩护。这些富人能够合理地意愿“他们继续现在的行动，以及每个人都相信其行动在道德上是可辩护的”为真。

在讨论康德的普遍法则公式时，科丝嘉写道：

这种观点拟定围绕的及其最有效地处理的是如下情形：使自己成为例外的诱惑、自私、吝啬、利用他人，以及漠视他人的权利。正是诸如此类之事而不是源自绝望或疾病的暴力犯罪，被康德用作不道德行为的典型。我认为，在这一点上他是无可挑剔的，因为诱使大多数人在其日常生活中犯下的那类恶正在于此而非别处。

我已经论证，康德的公式并没有最有效地处理自私、吝啬，以及利用他人。无论是自然法版本还是道德信念版本，康德的公式都没有谴责一些人利用另一些人的诸多做法——比如在男人、富人、强者利用女人、穷人、弱者的情

338

况下。而既然康德是将其公式作为道德的最高原则而提出的，那我们就能因其没有谴责此类行动而批评这个公式。诸如此类的自私和利用他人，恰好就是男人、富人和强者在其日常生活中受诱于且时常犯下的那类恶。

49 一个康德主义的解决方案

或许有人声称，在对康德的普遍法则公式提出这些反驳的过程中，我误读了该公式。内格尔提出，在我们问能否合理地意愿“每个人都奉行我们的准则”为真时，康德旨在让我们去想象，我们处于其他每个人的境地且在相关方面类似于所有此类的他人。这个提议使康德的公式类似于黄金规则的极度膨胀版，由此要求我们设法想象，自己将处于数以亿计的他人之境地。

在康德关于其公式的所有说法之中，没有任何一种会支持内格尔的解读。而且有相反的段落，比如康德对于自立的富人的讨论，这些富人持有“不帮助需要帮助的他人”的准则。康德声称，这个人不能合理地意愿其准则成为普遍法则，此时康德写道：

可能发生许多这样的情况，其中……由于来自其意愿的这样一条普遍法则，他就会剥夺自己得到帮助的一切指望，这种帮助是他自己想要的。

如果有意让这个人去想象自己将处于需要帮助的他人之境地，那康德就肯定会在这里讲出来。

339 内格尔用如下的主张来为其解读辩护：如果康德无意让我们去想象处于其他每个人的境地，那康德的公式就要经受严重的反驳。但即使最伟大的哲学家也可能忽视反驳。

罗尔斯对康德的公式提出了另一种解读。罗尔斯指出，在我们运用该公式时，康德意图让我们去想象，我们对自己或我们的处境一无所知。我们应该问的是，如果处在无知之幕（veil of ignorance）的背后，不知道自己的性别、贫富，是幸运的抑或需要帮助的，那么我们能够合理地意愿什么。与内格尔一样，罗尔斯用来支持其解读的主张是，似乎有必要替康德的公式辩护以使之免于反驳。但即使康德本该使用无知之幕的观念，那也没有显示他这样做过。在对其普遍法则公式的讨论中，康德从未提议，我们应当想象对自己或我们的处境一无所知。

根据威廉斯 (T. C. Williams) 对康德公式的第三种解读, 康德意图让我们从想象中的公正的观察者之视角来判断我们的准则。类似地, 威廉斯用来支持其解读的主张也是, 有必要替康德的公式辩护以使之免于反驳。但康德在讨论其公式时, 从未要求我们想象自己是公正的观察者。

斯坎伦提出了第四种解读。他提议说, 在我们运用康德的公式时, 康德意图让我们提问, 是否每个人都能够合理地意愿我们的准则成为普遍法则。但这不可能是康德的原意。康德写道:

我绝不应当行动, 除非以这样的方式行动: 我能够意愿我的准则同时成为普遍法则。

康德对其公式给出了许多不同的表述, 其中没有任何一种是指每个人都能够意愿什么。

如果不作为关于康德原意的主张, 而作为对康德公式的修正方法, 那么这些提议会更好, 由此就能避免我们一直在考察的这种反驳。

我认为, 在以上提出的修正之中, 斯坎伦的提议最佳。根据康德公式的道德信念版, 或者说: 340

MB: 按照某准则行动是不当的, 除非我们自己能够合理地意愿如下情况为真: 每个人都相信这样的行动在道德上是可允许的。

根据斯坎伦的提议, 这会变成:

MB4: 按照某准则行动是不当的, 除非每个人都能够合理地意愿如下情况为真: 每个人都相信这样的行动是道德上可允许的。

康德在关于其他两条原则即自主公式 (Formulas of Autonomy) 与目的王国的某些说法之中, 对这种修正也有所表示。例如, 康德指出:

这样一种观念, 即每个理性存在者的意志是作为给出普遍法则的意志。

尽管康德从未诉诸每个人能够合理地意愿什么，但这或许仅仅是因为他认为，对其公式的这种修正是多此一举。康德可能认为，任何一个人能够合理地意愿什么，必定等同于其他每个人能够合理地意愿什么。根据这一假定，MB与MB4总是吻合的。

我已讲过，这个假定是错误的。能够合理地许多男人、富人或强者所意愿之事，不能合理地许多女人、穷人或弱者所意愿。既然不同的人能够合理地意愿之事可能存在这样的差别，MB与MB4有时就会发生冲突，由此我们就必须在它们之间做出选择。康德如果看出有必要这么做，那原本就会正确地选择MB4。

接着回顾：我们应当修正康德的公式，以便它的适用对象不是行动者的准则，而是对此人在做之事的道德相关描述。由此我们的修正公式就可以变成：

MB5：按照某准则行动是不当的，除非每个人都能够合理地意愿“每个人都相信这样的行动是道德上可允许的”为真。

341

经过类似的修正，康德的自然法公式会变成：

LN5：按照某种方式行动是不当的，除非每个人都能够合理地意愿，每个人在类似的情形下都会尽可能地这么做。

如我在一个注释中所说明的，诉诸MB5就足够了。

在相信某种行动是道德上可允许的时，人们也就接受了某种允许此类行动的原则。因此，MB5可以变成：

普遍可意愿的原则公式 (the Formula of Universally Willable Principles)：某一行动是不当的，除非这样的行动为“每个人都能够合理地意愿该原则被普遍接受”的原则所允许。

用斯坎伦的话来说，“要回答关于正当与否的问题，我们必须问的是……‘我们所有人都能够意愿何种一般的行为原则？’”

该公式使我们得以进行不偏不倚的道德推理，其所采取的方式避免了稀罕性反驳、大赌注反驳，以及不可逆反驳。由于未诉诸行动者的准则，该公式避免了混合型准则反驳。由于允许我们诉诸有条件的原则，该公式也避免了门槛

反驳。我们还需要再做一个修正以避免新理想世界反驳，但这种修正会引起某些我们在此可予搁置的复杂性。

我曾说过，在考察了某些类似的反驳之后，有些人开始认为，康德的普遍法则公式不能帮助我们确定哪些做法是不当的。在用于此类问题时，伍德将该公式称作是“有根本缺陷的”和“近乎无价值的”；赫尔曼声称，不可能使之变得可行；希尔（Hill）怀疑它能否提供“即便是某种不严谨的和局部的行为指南”；而奥尼尔声称，它通常提供的指导要么不可接受，要么根本没有。这些主张是关于康德的*实际公式*，因而如我已论证的，它们是有正当理由的。某一行动是否不当并不取决于行动者的准则，因而康德的公式如果仅诉诸行动者能够合理地意愿什么，就不可能成功。但我们可以修正康德的公式，其途径是放弃康德诉诸的、在包含策略的意义上的准则概念，转而诉诸原则以及每个人都能够合理地意愿什么。由此所有这些反驳就不存在了。

342

如果我们所诉诸的、每个人都能够合理选择的原则是每个人都接受的，我们的观点就属于所谓契约论的（Contractualist）类别。若干作者比如罗尔斯和斯坎伦，提出了已被称作是契约论（Contractualism）的康德主义版本。但我认为，普遍可意愿的原则公式是最接近康德本人观点的契约论版本。那么，我们可以重申这个公式并给予一个更简洁的名称。根据

康德式契约论公式（the Kantian Contractualist Formula）：每个人都应当遵循这样的原则，其被普遍接受是每个人都能够合理地意愿的。

这个公式或许就是康德一直在努力探寻的东西：道德的最高原则。

第十五章 契约论

50 合理的协议公式

343 大多数契约论者都要求我们去想象，我们与他人拟就每个人都愿意接受的道德原则在努力达成协议。根据我们可称为的

合理的协议公式 (Rational Agreement Formula)：每个人都应当遵循这样的原则，其被普遍接受是每个人都会合理地同意的。

而有些契约论者诉诸的原则是，每个人都会合理地同意，该原则被普遍遵循或被成功地奉行。我的大多数主张都适用于诸如此类的契约论版本，这一点我将回头再讨论。我要讲的是，我们选择的原则是我们同意其被普遍接受的。大多数契约论者假定，如果我们的选择对于自己是最好的或预期最好的，我们就是在合理地选择。我们可以从这个假定入手。

尽管有些原则被普遍接受对于每个人都会是最好的，但还有其他一些原则被接受仅对于某些人是最的。例如，对于男人是最好的，并不总是对于女人最好。看来或许是，在人们的利益相冲突的情况之下，没有任何一个原则是每个人在自利的意义上都会合理选择的。但是，合理的协议公式仅适用于每个人都会合理选择的原则。那些其被接受对于我们自己会最好而为我们所选的原则，如果不能为其他某些人所合理地选择，那将是没有意义的。

344 我们能够合理选择什么，也将取决于未能达成协议所带来的影响。大多数契约论者建议我们去设想：如果我们没有达成协议，那将无人接受任何道德原则；由此无人会认为，有任何行动是不当的。这样一个世界对于每个人都可能

是不利的。在这种无道德的（*amoral*）无协议世界（*No-Agreement World*），如霍布斯（*Hobbes*）令人印象深刻地叙述的，我们的生命将是“孤独、贫穷、恶劣、残酷而短促的”。这会给予每个人以强有力的自理理由去努力达成协议。

我们可以设想，为了更易于达成协议，将展开讨论以及一系列假想的投票。但是，必须有某一最终的投票表决。我们所有人都必须明白，如果在最后一轮没能达成协议，我们就将失去最后的机会，因为我们不能再来一遍。在前几轮，我们策略性地投票将是合理的。我们可以宣告，我们意图选择并将投票给最有利于自己的原则，由此努力劝说他人也投票给这些原则。只有在决定性的最终投票中，鉴于达成协议的必要性，我们每一个人对他人做出充分的妥协才是合理的。

有些契约论者认为，道德最好被视为某种互利的协定（*bargain*）。这不一定是实际的协定。当人们的利益相冲突时，每个人就解决冲突的某些原则达成一致，这是合理的。通过诉诸这一事实，这些作者主张，我们可以在没有如此协定的现实世界证成这些原则。我们应当像我们本该合理地同意的那样来相互对待。这是一种讲得通的主张。

然而，要用这种方式来证成某些原则，我们就必须对如下主张做出辩护：每个人都本会就这些原则合理地达成协议。高契尔（*Gauthier*）在讨论他关于合理的协议公式的推荐版时，建议我们去“设想在每一方提出其最初的主张之后，协议就在单个回合的妥协之中达成”。但我们简直不可能设想，这样的协议将会达成。鉴于我们有必要达成协议，如下情况就是合理的：我们每一个人都设法预测他人会选择哪些原则，并就这些原则自行做出选择。在某些情况下，我们每一个人或许都能够预测其他人会选择什么。比如说，我们要就某组固定资源在我们之间的分配上达成协议。对于每个人来说唯一合理的选择或许是，每个人都应该获取平等的份额，因为我们能够各自预测，其他每个人将做出这一选择。但当我们在选择大多数其他道德原则的时候，这种协调（*coordination*）问题就不会有如此明显的解决方案。我们只能在暗中揣摩，设法预测他人会做何选择。因此，在决定性的最终投票中，就会有许多相冲突的原则，对于每个人来说选择它们是同等合理的。由此合理的协议公式就会失败，因为不存在任何一组原则是每个人都应当合理选择的。

契约论的这一版本还面临另一种反驳。无协议的世界对于某些人来说没那么不利，譬如那些更有能力的人，以及在非法的意义上控制着更多资源的富人。在无道德的世界，拥有诸如此类优势的人具备更强的谋生能力。每个人都

清楚，这些人没那么必要去达成契约论者的这种协议。这会给予他们更强的议价能力（bargaining power）。他们可能宣称，在决定性的最终投票中，他们将选择某些允许他们保持其优势并给予他们更多利益的原则。这样的威胁或许是可信的，因为与其他人相比，这些人更乐于承担产生无协议世界的风险。进一步说，在讨论某些问题的时候，无协议对于某些人来说可能更好。其中一例就是富人应当将其资源的多大份额转给穷人的问题。如果在此问题上没达成协议，乃至无人接受富人应当给出多少的任何原则，那将差不多等同于每个人都认为富人分毫不给是可允许的。这对富人来说或许不错。以如此种种方式，那些拥有更强议价能力的人或许能够运用这种能力，以使“他人接受有利于强者的原则”变成合理的。

346 有些作者接受合理的协议公式的这种蕴含。像高契尔这样的霍布斯式的（Hobbesian）契约论者就是如此，他们所辩护的仅仅是最低限度版本的道德（minimal version of morality）。高契尔宣称，既然道德预设着互利，那么在其他某些人的存在无益于我们的情况下，我们施大害于他们就没有错。比如说，在这种观点看来，当欧洲人在北美建立殖民地的时候，他们杀害土著居民在道德上是可允许的。对于关怀某些先天残疾者的要求，该观点也不能直接予以支持。高契尔承认，诸如此类的结论强烈地冲突于大多数人的道德信念。但高契尔拒绝诉诸这样的直觉信念或我们的“深思熟虑的道德判断”（considered moral judgments），他主张，道德理论应当置之不理。

我已拒绝高契尔的如下主张：当我们运用合理的协议公式时，高契尔的最低度道德是每个人都应当合理选择的。如我在附录B所论证的，我们还应当拒绝高契尔关于合理性的假设。而且我认为，我们应当拒绝高契尔的道德观点。与洛克（Locke）对霍布斯的评论一样，高契尔的最低度道德不承认“大量显而易见的义务”。我认为，类似的断言也适用于其他霍布斯式的理论。霍布斯式契约论者的论证是不健全的，其结论是不可接受的。

51 罗尔斯式契约论

罗尔斯尽管也诉诸合理的协议公式，但其所辩护的结论更可接受。罗尔斯的大多数主张，都是关于他所称的、在民族国家（nation-states）的社会之中的基本结构（basic structure）或主要制度的正义（justice）。这些主张在此是不相干的。我的评论将仅仅涉及罗尔斯对道德的契约论解释，即，他所称的

作为公平的正当性 (rightness as fairness)。

我将论证，罗尔斯版的契约论在运用于道德时是失败的。而如果我们把契约论从罗尔斯的大作《正义论》中移除，那将得到一种自由的平等主义观点；该观点不仅本身极具吸引力，而且也得到罗尔斯的某些非契约论主张和论证的充分支持。

我们对罗尔斯的道德契约论的考察，可以从他关于合理性和理由的假设入手。罗尔斯接受一种基于欲望的主观理论，他主张我们理性上应该努力实现这样的目标，即我们在完全知情 (fully informed) 和程序上合理的慎思 (deliberation) 之后最想要的那些目标。接受这种理论的许多人都相信它与合理利己主义相一致，后者主张，我们应当合理地努力去做最有利于己的任何事。这些人错误地假定：经过这样的慎思，我们每一个人总是最关心自己余生的总体福祉。

罗尔斯并没有做出这样的假定。在他所考虑的情形之中，正义要求我们按照不利于我们的方式行动。罗尔斯主张，即便在这样的情况下，我们去做正义要求之事也可能是合理的。我们如果做经全面考虑之后最想做的事，那将是在合理地行动。用他的话来说：

如果一个具有慎思理性 (deliberative rationality) 的人想要从正义高于其他一切的立场行动，那么这样的行动对他来说就是合理的。

然而，既然罗尔斯的理由理论是基于欲望的，他就无法断言，正义地行动对于每个人都将是合理的。罗尔斯在讨论这些人——他们如果不正义地行动，其明智的欲望将更好地得到满足——的时候宣称，他们不会有充足的理由去履行正义的要求。

如我已论述的，按照主观理论，我们不可能有理由把某事物作为目的或由于该事物自身之故而想要它。如果人们不关心某事物，即便经过完全知情和程序上合理的慎思之后仍不关心，那我们就不能宣称他们有理由关心。罗尔斯会接受这些主张。他也写道：

我们知道人们是理性的，但不知道他们将要追求的目的，只知道他们会明智 (intelligently) 地去追求。

与此类似，罗尔斯在讨论如下观点时，

某件事是正当的……一旦某位完美理性而公正的旁观者会予以赞许；

他写道：

由于这一界定没有对公正的旁观者做出任何具体的心理假设，那得不到任何原则来解释他的赞许……

在此罗尔斯认为，我们没有关心任何事的任何理由。如果认为我们有这样的理由，那他就不会断言：我们如果所知道的仅仅是某人是完美理性的（*ideally rational*），那就不能对这个人会赞许什么得出任何结论；而是会这样断言：这个人既然是完美理性的，那就会赞许他最有理由赞许之事。例如，他会赞许减轻痛苦或挽救人们生命的行动。

作为契约论者，罗尔斯诉诸这样一种原则：如果我们所有人就全体接受的原则在努力达成一致，那么该原则就是每个人都会合理选择的。按照罗尔斯基于欲望的理论，人们合理的选择是什么，取决于他们实际上想要什么。由于无法预测人们想要什么，罗尔斯就补充了一个动机假设（*motivational assumption*）。他建议我们去设想，当我们在选择道德原则时，每个人的主要目标将是促进其自身利益。按照这一假设，罗尔斯基于欲望的理论就与合理利己主义相一致。如果我们最关心自身利益，那么根据基于欲望或目标的理论，做出预期最能促进这些利益的选择对我们来说就是合理的。由此罗尔斯的动机假设就让他诉诸关于自利的合理性的主张。用他的话来说，

在选择原则时，每一个人都尽最大可能地努力促进自己的利益。

通过添加无知之幕，罗尔斯修正了合理的协议公式。根据

349 罗尔斯的公式（*Rawls' Formula*）：每个人都应当遵循这样的原则：如果每个人都必须在不知道关于自己及其处境的任何特定事实的情况下达成协议，那么他们就都会同意该原则被普遍接受，这在自利的意义上是合理的。

在解释为什么添加无知之幕的过程中，罗尔斯诉诸前面提及的、对霍布斯式契约论的两个反驳。

第一，如果每个人都知道关于自己及其处境的特定事实——比如其性别、年龄、能力，以及他们所控制的资源——我们就不可能指望解答这一问题，即，选择什么对于每个人来说会是合理的。用罗尔斯的话来说，“议价问题……将复杂得令人绝望”。不存在任何这样的原则，其被普遍接受是每个人都会合理地同意的。罗尔斯的无知之幕解决了这个问题。如果无人知道任何有关自己如何相别于他人的事实，那么每个人都选择同样的原则就是合理的，由此可确保达成协议。“选择什么对于任何一个人来说会是合理的”这样的提问就够了，因为同一个答案将适用于每个人。

第二，如罗尔斯所指出的，如果我们对自己及其处境一无所知，那将使我们变得不偏不倚。我们不会知道那些可能给予我们以更强议价能力的事实。任何人也不可能选择偏袒其自身利益的原则。尽管我们会出于自利的理由去选择原则，但我们的无知将确保在选择原则的过程中，我们给予每个人的福祉以同等的权重。

罗尔斯写道，其主要目标之一是提出一种系统的理论，以替代各种形式的效用主义。令人惊讶的是，在努力实现这一目标的过程中，罗尔斯提出的道德契约论版本诉诸自利的合理性与不偏不倚的某种结合。我们应该预期，这样一种理论会支持某种属于或近似于效用主义的观点。如罗尔斯本人所指出的，大体上说，效用主义是自利的合理性加上不偏不倚。

罗尔斯意识到这个问题。根据罗尔斯公式的一个版本，在想象我们在无知之幕背后的时候，我们会假定，我们处在任何人之境地的机率均等。罗尔斯声称，按照这个假定，选择其被接受将产生最大可能的平均福祉（average level of well-being）的原则，对于每个人来说就是合理的。通过选择这一平均效用主义原则（Utilitarian Average Principle），我们每一个人都使自身的预期福祉最大化。

350

我们可将此称作机率均等公式（Equal Chance Formula），罗尔斯拒绝该公式。罗尔斯宣称，我们如果处在无知之幕的背后，就不应当假定我们处在任何人之境地的机率均等。根据罗尔斯对其公式的偏好版（preferred version）——我们可称之为无知公式（No knowledge Formula），我们将不知道概率（probabilities）。罗尔斯论证说，这将使得选择某种非效用原则对我们来说是理性的。

罗尔斯的契约理论要实现其目标就必须证明，他对机率均等原则的拒绝是

正当的。在描述无知之幕的时候，罗尔斯写道：

似乎没有任何客观的根据……来支持这一假定：一个人有变成任何一人的均等机率。

这一评论将我们处在无知之幕背后的想象状态，看做仿佛是某种实际的事态，其性质是我们必须接受的。但是，罗尔斯是在提出一项思想实验，他可以自由选择实验的细节。他可以建议我们去设想，我们有成为任何人的均等机率。因此，罗尔斯必须给出对机率均等公式的其他反驳。罗尔斯本人指出，既然存在着具有不同蕴含的多种契约论公式，他就必须替他所选择的特定公式作辩护。他写道，他的公式必定是那种由于“最好地表达了加于原则选择的条件，这样的条件被人们广泛认为是合情理的”，从而“在哲学上是最优的”。与机率均等公式相比，其无知公式更好地表达了这些条件——罗尔斯能够这样宣称吗？

351 我认为答案是否定的。罗尔斯的无知之幕旨在确保，我们在选择原则的过程中是不偏不倚的。要实现这一目标，罗尔斯没有必要要求我们假定对概率一无所知。如果我们设想处在任何人之境地的机率均等，那将使我们变得同样不偏不倚。既然机率均等公式与无知公式没有任何其他的差别，那就无法宣称，罗尔斯的无知公式就其自身而言更有道理。

在讨论其所称的、对其理论的“康德主义解读”时，罗尔斯提出了对其无知公式的又一种辩护。罗尔斯写道，康德式契约论

旨在支持尽可能最厚的（thickest）无知之幕……康德主义的理据……始于让各方毫不知情，然后添加刚好足以让他们能够形成合理协议的信息。

罗尔斯提出，通过假定我们的所知尽可能少，我们就会让我们的道德推理，尽可能类似于在康德的永恒本体界之中我们的本体自我所做的推理，而我们由此就最好地表达了我们的自由和自主。

我认为，对无知公式的这一辩护并不成功。我们如果始于这样的假设：处在罗尔斯的无知之幕的背后我们没有任何信息，而接着应当添加信息，使之刚好足以可能形成合理的协议，那就应当诉诸无知公式的一种更极端版本。比

如说，在做选择的过程中，我们没必要知道，不同的人有不同的能力或我们生活在一个资源匮乏的世界。即使不知道这些事实，我们所知道的也足以做出合理的决定。由此我们就会更接近于实现罗尔斯“尽可能最厚的无知之幕”的目标。但不能宣称契约论的这种版本是如罗尔斯所说的，“最好地表达了加于原则选择的条件，这样的条件被人们广泛认为是合情理的”。我们不可能合情理地要求，那些在选择道德原则的人尽可能地无知。我们能够更有理由诉诸的，是信息充分的（well-informed）而不是信息欠缺的（ill-informed）选择。罗尔斯也写道，根据其观点的康德版，“我们开始根本就没有任何信息；因为康德的消极自由（negative freedom）意味着，能够不受异己（alien）原因的決定而独立地行动”。真信念不能被恰当地视为异己的原因。

继而回顾罗尔斯的断言，机率均等公式“自然而然地导致”平均效用主义原则。既然罗尔斯无法就拒绝其契约论的机率均等版提出正当的理由，罗尔斯的理论就没有如他所期望的那样，提供了一种反对所有形式的效用主义的论证。

罗尔斯可能会回应说，我们还可以有另一种理由去拒绝某些公式或道德理论。如果某一公式的蕴含与我们某些最深思熟虑的和最坚定的道德信念之间发生太大的冲突，那么无论该公式看上去如何可行，我们都可以有正当理由予以拒绝。罗尔斯认为，效用主义与某些这样的信念相冲突，比如说，奴隶制总是不当的。他或许由此宣称，我们可以正当地拒绝机率均等公式，其根据是该公式由于导致平均效用主义原则而具有不可接受的蕴含。

但即使罗尔斯这么讲，其契约论仍然未能提供反对效用主义的任何论证。为了证成“拒绝机率均等公式而诉诸无知公式是正当的”，罗尔斯将诉诸我们的非效用主义（non-Utilitarian）信念。因此，他不能同时这样宣称：通过拒绝机率均等公式而诉诸其无知公式，我们能够证成我们的非效用主义信念是正当的。我们如果仅通过诉诸某些信念来为某一论证辩护，那就不能接着再通过诉诸这一论证来为这些信念辩护。这样的辩护以假定其所要设法证成的东西为前提而陷入了循环论证。

接着罗尔斯或许退而声称，尽管机率均等公式支持效用主义，但其无知公式会支持可行的非效用主义原则。如果这一点是真的，那么罗尔斯诉诸其公式，这至少表明无知之幕的契约论者（Veil of Ignorance Contractualists）不必接受效用主义的结论。

然而，罗尔斯的公式没有支持可行的非效用主义原则。在运用其公式时，罗尔斯争辩说，如果不知道概率，我们就应当合理地设想最不利的情况，且设

法使我们最不利的可能结果变得尽可能地有利。由此我们就应当选择这样的原则，即接受该原则将使最不利者的状况变得尽可能地有利。由于这一论证建议我们将福祉的最小值最大化，我们可称之为最大的最小值论证（Maximin Argument）。

353 这个论证受到了广泛的批评。而它即便是合理的，也不会支持某种可接受的非效用主义的道德观点。先假设在治疗都在患病的各种年轻人时，我们必须确定如何运用某些稀缺的医疗资源。有两种可能结果，其一是

布鲁将活到25岁，而1000名其他人都将活到80岁。

另一种结果是

布鲁将活到26岁，而其他这些人都将活到30岁。

接着我们可以设想，人们如果寿命短些，相对来说就更不利。根据最大的最小值论证，那我们就应当选择第二种结果，即让布鲁多活一年，因为这是对最不利者来说最好的结果。这是个无法得到辩护的结论。尽管我们可以有理由让那些最不利者优先受益，但这种优先性不应该是绝对的。这样做是不当的：选择让布鲁多活一年，而不是让其他1000人多活50年——这些人如果没有多活的这些年数，将全都和布鲁差不多地早死。在运用于诸多与此类似的情况时，最大的最小值论证具有未免太过极端的蕴含。

罗尔斯接受我的以上主张。尽管罗尔斯是将其最大的最小值论证运用于社会基本结构，但他同意，一旦我们将之运用于有关分配正义的其他问题，这个论证的蕴含太过极端。他认为，效用主义未能提供某种可接受的、关于分配正义的一般原则。但罗尔斯承认，其契约论版本也未能提供这样的原则。

现在我们可以转向其他的道德问题。根据罗尔斯的最大的最小值论证，在不同的道德原则之间进行选择时，我们应当合理地选择那些其被接受将最有利于最不利者的原则。最大的最小值论证即便是合理的，也无法可行地应用于许多道德问题。假设就“我们何时能够有正当理由不守诺、撒谎、或将风险强加于他人”问题，我们在比较不同的相关原则。那将很难确定，关于这些问题的哪些原则被接受会最有利于最不利者。这也不可能成为在这些原则之间进行选择的正当方式。设想有这样的情况：我们如果都接受许诺实践的两种方式之一，或者

354

是关于加风险于他人的两种原则之一，那就会给大多数人带来大得多的利益。这样的事实将不会如最大的最小值论证所蕴含的那样是道德上不适当的。

即使罗尔斯未诉诸这一论证，其公式在另一种意义上也没能支持可行的非效用主义原则。契约论的罗尔斯版本迫使我们忽略大多数非效用主义考虑。按照效用主义，在行动或原则之间进行选择时，我们知道产生的利弊之大小和数量就够了。我们大多数人相信，还存在若干其他具有道德重要性的事实与考量。例如，我们拥有关于利弊应该如何在不同的人之间分配的信念，拥有关于责任、应得、欺骗、强制、公平、感激以及自主的信念。在我们运用罗尔斯版的契约论的时候，诸如此类的一切考虑除非影响到我们自身的福祉，就都变得不相干了。尽管罗尔斯式的道德推理不同于效用主义推理，但只是减法（**subtraction**）上的区别。在描述人们如何从无知之幕后选择道德原则时，罗尔斯写道，他们

决定的唯一根据是，预计什么东西似乎能最好地促进他们能够确定的利益。

罗尔斯只是拒绝给予这些人以自利计算所需的大多数知识。既然罗尔斯想象中的立约者是出于纯粹自利的理由来选择原则，那就绝无可能纳入非效用主义的考虑。

罗尔斯在首次提出其理论时写道：

完全可能的是……某种效用原则得到采用，由此契约理论最终导致对效用主义的某种更深入而又更迂回的证成。

他还写道：

对于契约观——它在传统上是效用主义的替代者——来说，这样的结论将是一场灾难。

罗尔斯或许可以否认，其契约论版本能够证成某种形式的效用主义。但他的主张不得不是这样的：即使其理论导致某种效用主义的结论，但这不足以可行地证成这个结论。

52 康德式契约论

为了形成更可行、更成功版本的契约论，我们应该回到不同的公式，以及关于理由和合理性的不同观点。根据

康德式契约论公式：每个人都应当遵循这样的原则，其被普遍接受是每个人都能够合理地意愿或选择的。

接着回顾，根据

合理的协议公式：每个人都应当遵循这样的原则，其被普遍接受是每个人都会合理同意的。

这两个公式都要求一致同意，因为它们诉诸的原则都是其被普遍接受是每个人都能够合理选择的。但与合理的协议公式不同的是，康德式公式没有使用协议的观念。在运用协议公式时，我们想象我们全都就哪些原则会为每个人都接受的问题在努力达成协议。这样的协议是必要的，因为每个人所愿意接受的，都只是在这单次的思想实验之中每个人都选择的原则。相比之下，根据康德的公式，

如果每一个人都各自认为，每个人都会接受他或她本人选择的原则，那么每个人都应当遵循每个人都能够合理地选择的原则。

356 在运用这一公式的过程中，我们进行多次思想实验，一人一次。在这些分别进行的选择之中，我们无人有必要与他人达成协议，因为我们每一个人都拥有这样的能力去选择哪些原则会为每个人都愿意接受。康德式公式以一种相当不同的方式来要求一致同意。该公式诉诸的原则是在多次的、分别的思想实验之中，每个人都会有充足的理由去选择的。

尽管罗尔斯正确地拒绝了合理的协议公式，但我认为，康德式公式比罗尔斯的公式更可行，且能更好地实现罗尔斯的目标。

罗尔斯的无知之幕，其部分旨意在于消除议价能力上的不平等。康德式公式以更好的方式实现了这一目标。既然没有达成协议的必要，也就没有议价的机会，因而无人会具有更强的议价能力。在问每个人都能够合理地选择哪些原则时，我们由此就可以假设，每个人都知道一切相关的、给予理由的事实，进而能够对所有这些理由予以回应。

接下来考察罗尔斯拒绝效用主义的一个理由。效用主义认为，给少数人施加重负的做法，只要会给其他人带来更大量的利益就是正当的。罗尔斯主张，在此情况下，正义

不允许以强加于少数人的牺牲为代价，来换取多数人享有更大量的利益。

根据若干作者的观点，效用主义者之所以得出如此不可接受的结论，是因为他们将不同人的利益和负担简单相加。用内格尔的说法，不同人的要求被全都“扔进料斗”，然后混合成一个冷冰冰的数目。其中有些作者提议，为了保护人们免于被施以这样的重负，我们应该转而诉诸全体一致的协议（*unanimous agreement*）的观念。根据该提议，通过要求这种协议，我们给予每个人以不被迫承受此种负担的否决权，由此实现我们可称为的反效用主义的保护性目标（*anti-utilitarian protective aim*）。

然而，否决权可能遭到误用。恰恰是由于要求这样的全体一致的协议，合理的协议公式更难达成这一保护性目标。这种公式给出了更进一步的有利条件，但其所给予的对象不是最需要而是最不需要道德保护的那些人，因为后者更强的能力或控制的更多资源给予他们以更强的议价能力。

357

罗尔斯的公式在实现这种保护性目标方面鲜有作为。其无知之幕尽管消除了议价能力上的差别，但也使得任何人都不知道，他们是否属于某些效用原则会要求或允许我们施以重担的少数人。罗尔斯所诉诸的原则，其被选是在自利的意义上合理的。而如我之所论，罗尔斯没有对机率均等公式提出恰当的反驳。因此，他就无法言之有理地否认：在无知之幕之下，我们能够合理地选择某种效用主义原则，或者是某种类似的但或多或少更谨慎一些的原则，由此为了获得极有可能的收益而承受小概率的大损失之风险。

康德式公式提出了一致同意的要求，但既未诉诸无知之幕也无需达成协议。部分地是由于这一原因，该公式更好地实现了保护性目标。效用主义者如

果诉诸该公式就不得不主张：我们即使知道我们属于将会被施以重担的少数人，也能够合理地选择效用主义原则。至少在某些情况下，我们能够言之有理地拒绝这种主张。

康德式公式还有其他优点。罗尔斯的无知之幕尽管确保了不偏不倚，但其做法很粗糙，如同额叶切除术。不同人之间的分歧并未解决，而是被压制。无人知道关于自己及其处境的任何信息，因而可以保证一致同意。在康德式公式所诉诸的思想实验之中没有无知之幕。每个人都知道自己的利益如何冲突于他人的利益。既然未保证一致同意，那么若能达成一致同意则更具道德意义，因为存在某些即使在信息充分的情况下每个人都都能够合理地同意的原则。

是否存在这样的原则，取决于我们应当持有何种关于理由和合理性的信念。如果最优理论要么是合理利己主义，要么是某种基于欲望的或基于目标的主观理论，那康德式公式就不会成功。在该公式所诉诸的思想实验之中，不会产生任何一组这样的原则：其被选是每个人在自利的意义上合理的；也不会产生一组这样的原则：其被普遍接受会最大地实现每个人在充分知情的情况下的欲望或目标。

我认为，我们应当拒绝所有的主观理论。尽管合理利己主义就其为客观的与基于价值的而言是一种正确的理论类型，但过于狭隘。根据我认为最正确或最优的那种客观理论，我们具有强有力的理由去关心自身的福祉，且是以一种时间上中立的方式去关心。但我们自己的福祉并非如合理利己主义者所主张的那样，是唯一最合理的终极目标。我们能够理性地同样看重其他的事物，比如他人的福祉。

再回到这个事实：由于罗尔斯诉诸的原则是出于自利的理由而被合理选择的，我们在运用罗尔斯的公式时就根本无法纳入非效用主义的考虑。我们在运用康德式公式时可以诉诸各种非道义式理由，因而该公式就能支持非效用主义的原则。

康德式公式要能成功，就必须充分满足某种我们可称之为的唯一性条件（**uniqueness condition**）。情况必须是这样的：至少在大多数情形下，有且仅有某种相关原则是每个人都能够合理地选择的。如果没有这样的原则，那就不会有康德式公式要求我们去遵循的任何原则。那么，这个公式就可能由于未能拒绝显然不当的行动而失败。如果每个人都能够合理地选择两种及以上有着严重冲突的原则，那该公式就会以类似但更为复杂的方式而又告失败。但如果是每个人都能够合理地选择几种类似原则之中的任一种，那就是没有关系的。

这样一些原则将会是某种更一般的高阶原则（higher-level principle）的不同版本，由此在这些低阶原则（lower-level principles）之间的选择能够以某种别的方式来进行。我认为，唯一性条件通常可以得到充分满足。

为了说明康德式公式，我们可以将之运用于一个简单的问题。设想：

一定量的无主益品可为不同的人分享；

没有任何人对这些益品具有特殊的要求权，比如说基于他们具有更大的需要、生产了这些益品、或比其他人过得更差；

以及

若这些益品被均等地分配，则产生的利益最大。

在此类例子中，每个人显然应该被给予均等的份额。康德主义者可能论证道：

（A）每个人都能够合理地选择这样的原则：它在此类例子中给予每个人以均等的份额。

（B）无人能够合理地选择这样的原则：它给予自己及某群体中的其他人少于平均数的份额。

（C）唯有均等份额的原则不会给予任何人少于平均数的份额。

因此

（D）这是每个人都能够合理地选择的唯一原则。

如果接受合理利己主义，那我们就必须拒绝这一论证的第一个前提。根据该理论，每个人应当合理地选择某个给予自己多于平均份额的原则。如果接受某种关于理由的主观理论，我们也必须拒绝（A）。有许多人如果选择均等份额的原则，那么他们在充分知情的情况下的欲望或目标不会得到最大的满足。但我认为，如（A）所主张的，每个人都能够合理地选择均等份额的原则，因为我们所有人都有充足的理由来做出这一选择。我们不会被合理地要求去选择某种这样的原则，即它给予我们的份额多过平均数。如（B）所主张的，无人

能够合理地选择任何这样的原则：它给予自己及某群体中其他人的份额少于平均数，由此产生较少的利益，且被不平等地分配。如（C）所主张的，唯有均等份额原则给予任何人的份额都不少于平均份额。因此，如这一论证所表明的，这是每个人都能够合理地选择的唯一原则。康德式公式正确地蕴含着：在此类情况下，每个人都应该得到均等的份额。

53 斯坎伦式契约论

现在我们可以介绍契约论的另一版本。根据

斯坎伦的公式（Scanlon's Formula）：每个人都应当遵循无人能够合情理地予以拒绝的原则。

其更完整的陈述是：

恰好当（just when）某个行动为某个无人能合情理地拒绝的原则所禁止，或者当允许它的任何原则都可被至少一人合情理地拒绝时，该行动就是不当的。

尽管“合情理的”有时与“合理的”同义，但斯坎伦的公式是在另一种意义即部分是道德的意义（partly moral sense）上使用该词。在此意义上，我们如果给予他人的福祉或道德要求以过小的权重，那就是不合情理的。

有些人宣称，由于斯坎伦诉诸这种部分是道德的意义上的“合情理的”，其公式是空洞的。他们说，如果我们接受斯坎伦的公式，那将对我们的道德思考毫无助益，因为每个人都可以宣称，其所接受的道德原则不能合情理地予以拒绝。

这种反驳忽略了如下事实：在运用某些契约论公式时，我们不能诉诸关于哪些行动是不当的信念。再设想

361

手段情形（Means）：格蕾和怀特被困于正逐渐坍塌的废墟之中。格蕾毫无危险。我可以救怀特的命，但只有以格蕾的身体做掩护才行，这种方式未经格蕾的同意且会毁掉她的腿。

许多人都会认为，我用这种方式去救怀特的命是不当的。我们如果接受这一观点，就可能诉诸

有害手段原则（the Harmful Means Principle）：给某人施加像这样的严重伤害以作为让其他人受益的手段，这是不当的。

根据另一种与之相冲突的观点，我们可称之为

更大负担原则（the Greater Burden Principle）：给某人施加一个负担如果是使其他某人免于大得多的负担之唯一途径，那就是可允许的。

关于拒绝道德原则的合情理的依据是什么，斯坎伦提出了各种主张。根据其中之一，

在各种替代原则都会给他人施加大得多的负担的情况之下，某一原则会给你施加小得多的负担，你因此拒绝该原则……这是不合情理的。

按照斯坎伦的预定涵义，我们在未能给予某人以某种利益时就给这个人施加了一个负担。怀特可以争辩说，如斯坎伦的主张所蕴含的，格蕾不能合情理地拒绝更大负担原则。尽管我按该原则行动会给格蕾施加一个负担，但我按有害手段原则行动将给怀特施加一个大得多的负担。与未能保全生命相比，失去一条腿的负担要小得多。

格蕾或许答道，按照她的观点，怀特不能合情理地拒绝有害手段原则。但这种拒绝为什么就是不合情理的？格蕾或许说，她有这样的权利：在未经本人同意的情况下，她不应作为有利于他人的手段而受到严重伤害。但在对拥有这种权利的宣称之中，格蕾就在隐含地诉诸她的如下信念：我以这种方式来伤害她是不当的。在运用斯坎伦的公式时，我们不能诉诸像这样的道义式信念。格蕾可能宣称：

（1）我的行动是不当的，因为无人能够合情理地拒绝禁止这种行动的有害手段原则。

但格蕾不可能用如下主张为（1）辩护：

（2）无人能够合情理地拒绝这一原则，因为这样的行动是不当的。

如我所言，如果将这两个主张合并，我们会陷入循环论证，一无所成。格蕾必须用某种其他的方式去论证“无人能够合情理地拒绝有害手段原则”。

如这个例子所示，斯坎伦式契约论远非空洞的。在拒绝有害手段原则时，怀特能够诉诸这一事实：与失去一条腿相比，死亡的负担要大得多。根据斯坎伦的观点，这是一种可以为拒绝某道德原则而提供合情理根据的事实。在辩护有害手段原则时，格蕾不可能诉诸任何诸如此类的事实。格蕾的难题在于，不同于更大负担原则，对有害手段原则的最佳辩护是通过诉诸我们关于哪些行动是不当的直觉信念。我们许多人会相信，未经本人同意而给某人施加严重伤害是不当的，即使在这是救他人生命的唯一途径时也是如此。但在运用契约论公式时，我们不能诉诸这样的信念。

与罗尔斯一样，斯坎伦在一定程度上是作为避免行为效用主义（Act Utilitarianism）或AU的一条途径而提出其契约论的。如我们刚刚看到的，在某种意义上，契约论使AU更容易得到辩护。我们大多数人之所以拒绝AU，是因为该观点要求或允许众多在我们看来是不当的行动。如斯坎伦之所述，

行为效用主义的蕴含与根深蒂固的道德信念之间存在极大的分歧。

但在运用某些契约论者的公式且遵循道义式信念的限制时，我们不能诉诸
363 上述信念。

但是，即使不诉诸这样的信念，斯坎伦式契约论者也能够拒绝行为效用主义。为了说明斯坎伦的公式，有些例子是值得考察的。设想

移植情形（Transplant）：我在医院接受某个小手术。你是我的医生。你知道，如果秘密地将我杀掉，那么我被移植的器官会用来挽救其他五个人的生命。

根据

AU：我们应当总是做或努力做最有利于人们的任何事。

这一原则要求你杀死我，去救这五个人的生命，因为这是你最有利于人们的做法。我们大多数人会认为，这种行动是不当的。

通过诉诸斯坎伦公式的一个版本，我们可以言之成理地替这个信念辩护。设想我们全都知道，只要我们在医院就医，我们的医生就可能秘密地杀死我们，以便我们的器官可用来挽救其他人的生命。即便这种风险极低，对此的了解也会使我们许多人感到焦虑，且会恶化与我们的医生之间的关系。医患关系极为重要，因为我们通常依赖于医生的诊断、他们对我们福祉的关心，而且他们或许是我们能指望帮助我们脱离生命危险的人。通过诉诸此类事实，我们能够合情合理地拒绝AU。如果在这样的情况下所有医生都遵循上述原则，那能被救活的患者要多一些。但是，多救的这些人之生命，其价值尚不足抵消如下方式带来的危害：如果我们所有人都知道，我们的医生认为用这种方式将我们秘密杀害是正当的，那对我们及他人来说是很糟糕的。我们可称之为焦虑和不信任论证（Anxiety and Mistrust Argument）。

这一论证以另一种方式说明：如果我们诉诸某种契约论公式，那将有益于我们的道德推理。如果孤立地考察移植情形，我们可以忽略这一论证。既然你可以通过秘密杀害我而救活那五个人，那你的行动就不会产生焦虑和不信任。但是，在运用某些契约论公式比如说康德式公式或斯坎伦式公式时，我们不会孤立地考察特定的行动。我们问的是，如果我们在选择每个人都愿意接受的原则，那么哪些原则是每个人都能够合理地选择的，或者是无人能够合情合理地拒绝的。就回答这个问题而言，我们必须考虑每个人都接受并知道每个人都接受这些原则而产生的效果。这就使得在移植情形中如下情况变得不相干：你的行动是秘密的，由此不会产生任何焦虑或不信任。

364

斯坎伦宣称，我们能够合情合理地拒绝某种原则，仅当能够提出某种更优的替代原则。如果拒绝AU，那我们应该提出何种替代原则？

将移植情形与另两个情形做个比较或许是有益的。再设想

隧道情形（Tunnel）：通过切换某铁轨的道岔，你可以让一列无人驾驶的失控火车变向，由此杀死了我而使另外五人幸免于死。

以及

桥梁情形（Bridge）：你可以救五人之命，其唯一方法是用遥控

使我跌倒于火车前，由此致我于死，但也启动了火车的自动刹车。

作为AU的替代，我们可以求助于

有害手段原则：施大害于某人以作为有益于其他人的手段，这是不当的。

按照这一观点，道德上重要的是：你救五人之命与你杀我这一行动之间如何发生因果关联。无论是在移植情形还是在桥梁情形中，你用杀我的方式去救五人之命是不当的。但在隧道情形中你杀我并非不当，因为在此你杀害我，将不是作为救五人之命的手段，而仅仅是作为让火车变向而产生的可预见的附带后果（side-effect）。我们许多人会接受这些主张，认为桥梁情形中的行动不当而隧道情形中的行动是可允许的。在运用斯坎伦的公式时，我们能够为这一区分做出言之有理的辩护吗？

365

我认为答案是否定的。在考察像隧道和桥梁这样的情形时，我们具有强理由去关注我们的生死，但绝无强理由去关注我们的死亡如何与救他人之命发生因果关联。我做出这一断言，并不是在假定重要的仅仅是结果。我们可以有理由去关心某些结果是如何产生的。但是，某个他人的可能行动方式是在杀了我们的同时救其他若干人之命，此时我们绝不会有这样的强理由：与其作为救这些人命的手段而被杀，不如作为其附带后果而被杀。在某种意义上，作为手段而被杀还好些，因为由此我们的死亡至少是在做善事。鉴于这些事实，斯坎伦的公式似乎不利于如下观点：在隧道情形与桥梁情形的行动之间存在某种重要的道德差别。如果我不能合情理地拒绝在隧道情形中允许你杀我的某种原则，那么这看来就是令人怀疑的——我能够合情理地拒绝在桥梁情形中允许你杀我的一切原则。斯坎伦的公式似乎蕴含着：这两种行动要么都不当，要么都是道德上可允许的。

接着考察对AU的另一种替代，它通过焦虑与不信任论证而提出。根据我们可称之为的

紧急原则（Emergency Principle）：医生必须绝不杀害病人以用作救更多人命的手段。但是，在某些非医疗的紧急情况下，每个人都允许去做救最多人命的任何事。

这种非医疗的紧急情况包括对人们生命的非故意而直接的威胁，比如说火灾、水灾、雪崩，或无人驾驶的失控火车。紧急原则谴责移植情形中你杀我去救五人，因为在此你是我的医生。但该原则允许你在隧道情形和桥梁情形中以某种方式杀我以救五人，因为这是非医疗的紧急情况，其中我对你来说是陌生人。

与有害手段原则相比，斯坎伦的公式似乎更强烈地支持紧急原则。该原则假定，道德上重要的，不是你救五人与你杀我之间的因果关联，而是在移植情形中你与我的私人（personal）关系，以及在医疗的与非医疗的紧急情况之间的其他区别。这些都是那种在运用斯坎伦的公式时我们可以更有理由诉诸的事实。我们有理由要求我们的医生相信，他们必须绝不杀害病人以用作救更多人命的手段——可以补充说，甚或附带后果。我们与医生之间的关系非常重要，而在上述罕见的非医疗紧急情况下，可能杀我们或救我们的那些人与我们之间并没有如此的私人关系。于是我们就有理由要求这些人相信，在此类情况下他们应当救尽可能多的人命。我们会认识到，如果我们的生命在诸如此类的紧急情况下受到威胁，我们更可能成为被救的一员。

54 道义式信念的限制

假设在努力思考这些想象中的情形之后我们相信：在隧道情形中作为救五人的附带后果，你杀我是道德上可允许的；但在桥梁情形中作为救人手段，你杀我则是不当的。那么，我们可能接受做出这种区别的有害手段原则。接着假设，由于我刚刚给出的理由，我们无法通过诉诸斯坎伦的公式来为有害原则辩护。对于如此种种原则的最佳辩护，是通过诉诸我们关于哪些行动为不当的直觉信念。但在运用契约论公式时，我们不能诉诸这些信念；在运用康德的普遍法则公式时也不能这么做。

现在我们可以挑战这一道义式信念的限制（Deontic Beliefs Restriction）。在诉诸这些康德主义的或契约论的公式时，我们为什么应该搁置关于哪些行动为不当的信念？

康德主义者和契约论者可能回答说，如果我们诉诸这样的道义式信念，他们的公式就是循环的，在某种意义上变得无用。如我所言，同时做出如下两点断言是无意义的：

当允许某些行动的任何原则都不能通过康德主义的或契约论的检

验时，这样的行动就是不当的；

367

以及

当且因为某些原则允许的行动是不当的，这些原则将不能通过上述检验。

但这并非令人满意的回答。这些公式会是无用的，除非我们遵循道义式信念的限制；但即便如此，这也并不表明，我们应当运用这些公式来进行道德思考。^①

另一种回答提出的主张是关于道德信念和主张的性质与可证成性（*justifiability*），在此意义上，它诉诸某种元伦理的（*meta-ethical*）区分。罗尔斯写道，根据直觉主义者（*Intuitionists*）的看法，存在某些关于哪些行动为不当以及哪些事实给予我们以理由的独立真理。奴隶制是不当的和我们有理由去预防或减轻痛苦，即为两种这样的真理。这些真理是独立的，意指它们并非出自我们的创造或建构。根据一种罗尔斯称之为建构主义（*Constructivism*）的不同观点，不存在这样的真理。在这种观点看来，孰对孰错完全取决于在某种康德主义或契约论的思想实验中，我们选择哪些原则是合理的。用罗尔斯的话来说，道德事实（*moral facts*）是什么，这是由我们来决定的。如果我们是建构主义的契约论者，且认为选择允许奴隶制的原则是合理的，那我们就应当断定奴隶制并非不当。尽管奴隶制可能在我们看来是不当的，但建构主义拒绝诉诸像这样的直觉道德信念，其中有些人认为这样的信念含有偏见、文化制约（*cultural conditioning*），或仅为幻象。

在此我姑且假设，我们应当拒绝这些怀疑论的（*sceptical*）、反直觉主义的观点。罗尔斯没有信奉建构主义，且经常假定存在某些诸如奴隶制为不当之类的独立真理。在设法实现罗尔斯所谓反思的平衡（*reflective equilibrium*）

① 尽管从上下文来看，“我们应当”改为“我们不应当”更易理解，但帕菲特在给译者的信中强调说，他此处的表达有点含糊，但用“我们应当”才能表达他的真实意思。他认为更清楚且更好的表达是这样的：“即使我们若不遵循这些限制，则这些公式就是无用的，但这也并不表明，若我们遵循这些限制，这些公式就是有用的，从而我们可以运用这些公式进行道德思考。”

时，我们应该诉诸我们的一切信念，其中包括关于某些类别的行动之不当性的直觉信念。如斯坎伦所言：

若作恰当的理解，这一方法是……对道德事务做出决断的最佳方式……确实说来，它是唯一可辩护的方法：取代它的显然是幻觉。

康德主义者和契约论者如果接受这一观点——我们的道德推理应诉诸这样的直觉信念，那就必须以某种别的方式为道义式信念的限制作辩护。 368

有一种直截了当且令人完全满意的辩护。要说明这种辩护，我们可以先区分两种意义：行动的某种属性和有关行动的某个事实可以在两种意义上使该行动成为不当的。行动的某种属性在使该行动为不当时，并没有因果地导致（cause）该行动为不当。在一种不重要的（trivial）意义上，不当性是作为非因果地（non-causally）使行动为不当的属性。其含义类似于蓝性质（blueness）是使事物为蓝的属性，非法性（illegality）是使行动为非法的属性。正是在某种与此不同的且极其重要的意义上，当行动具有某种其他的属性——比如说，造成无谓的痛苦之属性、许假诺的属性——的时候，这些事实可能非因果地使这些行动不当。造成无谓的痛苦不同于“是不当的”。但如果某行动造成无谓的痛苦，这一事实可能是由于使该行动具有“是不当的”不同属性而使之为不当。道德理论应努力描述在这种意义上使行动为不当的属性或事实。

斯坎伦曾经声称，其契约论给出了某种说明，这种说明不是关于什么使得行动为不当（what makes acts wrong），而是关于不当性本身，或者说是关于什么是行动之为不当（what it is for acts to be wrong）。我认为，这一说法是个失误。为解释其原因，我们可以先重述康德式契约论公式（KF）。根据

KF2: 某一行动是不当的，恰好当这样的行动为如下的原则之一所不容：每个人都能够合理地意愿该原则被普遍接受。

接着假定在

康德的意义，上，“不当的”意指“为其被普遍接受是每个人都能够合理地意愿的原则所不容”。

康德式契约论者如果是在这种意义上使用“不当的”，那就能够宣称是要对一种不当性给出说明。按照这一观点，在行动为这样的原则所禁止时，那就是在这种康德的意义上的什么是这些行动之为不当。但是，KF2由此就会成为一种隐含的重言式（concealed tautology），其公开形式之一是

KF3: 某一行动为这样的原则所不容，恰好当这样的行动为这样的原则所不容。

而这样的主张是不值得做出的。康德式契约论者应当转而以一种或一种以上的非康德的意义上使用“不当的”。KF2就不会由此变得微不足道，因为这样的断言意味着，一旦某一行动为这样的原则所不容，那就使得该行动在其他此类的意义上成为不当的。例如，康德式契约论者可能宣称：

KF4: 一旦某一行动为“每个人都能够合理地意愿其被普遍接受”的原则之一所不容（disallow），那就使得该行动在如下意义上成为不当的：它在他人面前无法得到辩护，值得谴责，以及是一种让行动者有理由感到悔恨而让他人有理由感到义愤的行动。

我们如果是康德式契约论者，就不应该宣称，我们的公式描述的是在其他这些意义上使得行动为不当的唯一属性或事实。存在着其他的、常常更具重要性的致不当（wrong-making）属性或事实。我们反而应该宣称，该公式描述的是某种高阶的致不当属性或事实，而一切其他此类的属性或事实均可归入或汇聚于它之下。例如，在某一行动是许假诺时，该事实或许使得它成为这样一种行动，即它为其被普遍接受是每个人都能够合理地意愿的原则之一所不容。根据这一版本的康德式契约论，这两种事实由此就都可以被真正宣称为使该行动为不当。

我认为，斯坎伦的理论应该呈现为同样的形式。根据

斯坎伦的公式：某一行动是不当的，恰好当这样的行动为无人能够合情理地拒绝的某个原则所不容。

斯坎伦如果在此是在另一种契约论意义上使用“不当的”，以意指“为这

样一种不可拒绝的原则所不容”，那他就可以宣称，其公式对不当性的这一契约论类型给出了说明，或者是对于什么是在这种意义上行动为不当的说明。但其公式由此就将成为又一种隐含的重言式，其公开形式之一会是这样的断言：某行动为这种不可拒绝的原则所不容，恰好当该行动为这样的原则所不容。无论我们的道德信念如何，我们所有人都可以接受该断言。斯坎伦的断言转而应该是：如果某行动为某种不能合情理地拒绝的原则所不容，那将使得该行动在一种及一种以上的非契约论意义上是不当的。

斯坎伦如今接受，其观点应该采取这种形式。我们由此可以说，按照斯坎伦的理论，在行动具有某些其他的属性时，那会使得这些行动为某种不可拒绝的原则所不容，从而这些事实都可以被真正宣称为，在其他意义上使得这些行动为不当。

契约论者如果做出这样的断言，那么就可以在无需将我们的道德直觉拒绝为无价值的同时为道义式信念的限制作辩护。按照契约论的这些版本，仅当我们在问契约论公式蕴含什么时，我们才不应该诉诸我们关于正在考察的行动之不当性的信念。当随后决定是否应当接受这些公式时，我们可以诉诸这些信念。比如在考虑关于哪些行动为不当的任何其他断言时，我们可以拒绝任何契约论公式——如果该公式的蕴含与我们直觉的道德信念之间存在过于频繁、过于强烈的冲突。

按照斯坎伦观点的这一版本，他并没有拒绝诉诸我们的直觉信念。相反，斯坎伦表明，我们不仅持有关于哪些做法为不当的此类信念，而且持有且能够有效诉诸关于何者为拒绝道德原则的合情理根据的直觉信念。此为斯坎伦对于我们的道德思考的最大贡献。

第十六章 后果论

55 后果论的理论

371 我们在问最优版本的契约论蕴含什么之前，回头考察什么是善的与什么是正当的之间的关系，这或许是有益的。

痛苦是我们有理由想要避免之事；有些人确实认为，痛苦在此意义上是坏的（bad）。但一些大哲学家没有这样的信念。比如说，休谟（Hume）并不是在蕴含理由的意义上使用“善的”或“坏的”。这或许就是休谟做出如下断言的原因所在：我们偏好公认较小而非较大的善，这不可能是不合情理的或违反理性的。休谟如果使用“较小的善”是意指“那些我们具有较少理由去偏好的东西”，那原本就不会认为这样的偏好是合情理的。休谟常用的“善”与“恶”，仅仅意指“快乐”与“痛苦”。

虽然在休谟看来，痛苦是恶的这种主张本是很平常的，但康德有时会拒绝这样的主张。例如，康德写道：

严格地说，善或恶适用于行为，而不适用于个人的感觉状态……因而有人可能会嘲笑这样的斯多葛学派哲学家，他在痛风的极度痛苦中高喊，“痛苦，无论你怎么折磨我，我都绝不会承认你是恶的（kakon, malum）”。然而，他是对的。

在主张痛苦不可能是恶时，康德的意思是痛苦不可能是道德上坏的。与休谟一样，康德有时似乎没意识到或忘记了，身处痛苦在蕴含理由的意义上是坏的。

372 罗斯（Ross）也是如此。罗斯认为，如果某事件是坏的，我们就有显见的

义务（*prima facie duty*）去尽可能防止该事件发生。由于我们没有这样的义务去防止自己处于痛苦状态，罗斯就断定，我们自身的痛苦并不是坏的。更确切地说，罗斯提出，我们的痛苦是坏的，但仅仅是从他人的视角来看是如此。罗斯之所以得出这一奇怪的结论，是因为忽略了在蕴含理由的意义上事物可有非道德的好坏之分。

痛苦不仅对于身处痛苦的人来说是坏的，而且是非个人地（*impersonally*）坏的。用内格尔的话来说，“受苦是件坏事，这毋庸赘言（*period*）；且不仅对于受苦者来说是如此”。许多人认为，虽然后果对于特定的人们来说可能有好坏之分，但后果可能有非个人的好坏之分这一说法是没有意义的。但如我所言，这样的意义可以得到解释。我们比较不同的可能后果，并且宣称某一后果会

是在蕴含不偏不倚理由的（*impartial-reason-implying*）意义上非个人地最好的，此时我们的意思是：从不偏不倚的视角看，这种后果是每个人都会有最大理由想要的，或者是希望会产生的。

在考察仅涉及和影响陌生人的可能事件时，我们的实际视角是不偏不倚的。但在我们的视角并非不偏不倚时，我们也有不偏不倚的理由，例如当我们能够减轻自己或他人的痛苦时就是如此。我们所有人都有理由为任何人——无论这个人与我们的关系如何——处于痛苦感到遗憾，在此意义上，一切痛苦都是坏的。而我们所有人都有理由希望每个人生活顺利。

我们如果接受某种理由的主观理论或合理利己主义，那就必须否认，后果可以有上述意义上的好坏之分。按照这些理论，绝不存在每个人都有理由想要或感到遗憾的后果。如果某个灾难或地震致众多人于死地，那也不可能在这种意义上是坏的，因为这一后果并非对每个人都坏，也并非每个人都有基于欲望或基于目标的理由而希望受难者不死。但如我已论证的，我们应当拒绝这些理论。

在下面的讨论中，我将在蕴含不偏不倚理由的意义上使用“最好”（*best*）这个词。通常有两种及两种以上的可能后果可被称作“同等最好”（*equal-best*）。鉴于这种说法会误导人们联想到精确性，因而还是称这样的后果不差于任何其他后果为好。但为简洁起见，我将用“最好”来指所有像这样的后果。

尽管任何讲得通的道德理论都可以诉诸关于后果的善性（*goodness*）之事

实，但某些理论将这样的事实视为根本的。根据我现在称作的

后果论：我们的行动正当与否，仅取决于有关事情如何进展会最好的事实。

后果论的理论可能产生多方面的分歧，因为它们如下两个问题上都可能提出相冲突的主张：什么是好与坏，以及行动的正当性（rightness）如何取决于有关“什么是最好的”事实。

有些后果论者是效用主义者，他们认为

（A）若人们在总体上得益最多，也就是说在除掉负担后他们得到了最大的利益，则事情就是进展最好的。

其他后果论者认为，后果的善性部分地取决于其他事实。例如，有人认为

（B）事情进展如何，部分地取决于利益与负担如何在不同的人之间进行分配。

按照两类这样的观点，两种后果中有一种即使包含的净收益较少，但利益与负担的分配更为平等，或最不利者得到的利益更多或负担更少，那么它可能是更好的。

“后果论”一词在某种意义上会误导人，后果的善性与使得事情进展最好这两种说法也是如此。这些用语让人产生这样的想法：在这些理论看来，要紧的仅在于将来以及我们行动的效果。后果论者可以拒绝这样的主张。某些后果的善性可能部分地取决于与过去相关的事实。例如，若按如下的做法则可能更好：将利益分给此前的较不利者；对死者守诺；只惩罚犯过罪的人。而某些做法、意图和动机，无论其效果如何，其本身可能就有好坏之分。比如说，善意的行动即便失败也可能是好的；而人们受骗或被强制，这通常就其本身而言就可能是坏的。在问若某事发生或某人以某种方式行动是否会最好时，我们问的是，从某种不偏不倚的视角看，每个人都最有理由去要求或希望什么。对于我们最有理由要求事情按照哪些方式来进展这一问题，这种意义上的“最好”完全未作解答。

然而，有一种价值观是后果论的理论所不能诉诸的。有些后果论者认为

(C) 在人们出于正当的理由而正当地行动时，这些行动本身就是好的，而不当的行动本身就是坏的。

如我在一个注释中所说明的，行动正当与否，不能取决于行动在这些意义上的好坏。但这对于这些理论并不构成严重的反驳。

所有的后果论者都诉诸关于什么使事情进展最好的主张。我们可将此称作后果论标准 (Consequentialist Criterion)。直接后果论者将此标准直接运用于一切事：不仅用于行动，而且用于规则、法律、习俗、欲望、情绪、信念、财富分配、大气状态，以及可能左右事情好坏的任何其他东西。有些人将此标准运用于行为，他们是行动后果论者 (Act Consequentialists)。其中有人主张

(D) 每个人都应当总是做实际上使事情进展最好的任何事。

另有人主张

(E) 每个人都应当总是做或努力做最可能使事情进展最好的任何事，或者更确切地说是使事情进展预期最好 (expectably-best) 的任何事。

如果 (D) 是在事实相对的 (fact-relative) 意义上而 (E) 是在证据相对的 (evidence-relative) 或信念相对的 (belief-relative) 意义上使用“应当”这个词，那么这两种主张并不冲突。在下面的大多数讨论中，我们可以忽略这两种主张之间的差别。而我通常会用“最好”意指“最好或预期最好”。

375

间接后果论者将后果论标准直接用于某些事，但只是间接地用于其他事。规则后果论者 (Rule Consequentialists) 将此标准直接用于规则或原则，但仅间接地用于行为。其中有些人认为

(F) 每个人都应当遵循这样的原则：其被普遍接受将使事情进展最好。

按照这种观点，尽管最好的原则是那种其被普遍接受会使事情进展最好的原则，但最好的或正当的行为并非会使事情进展最好的行为，而是为最好的原则所要求或允许的行为。会使事情进展最好的行动如果为最好的原则之一所不容，那就是不当的。与此类似，动机后果论者（Motive Consequentialists）主张，尽管最好的动机是那种其被每个人持有将使事情进展最好的动机，但最好的或正当的行动并非会使事情进展最好的行动，而是为拥有最好动机的人所做出的行动。这些理论与那些成理论体系形式的美德伦理（virtue ethics）之间有交集，后者诉诸最能促进人类繁荣或福祉的性格特征或其他气质。还可能存在许多其他形式的间接后果论。

56 后果论的准则

有些后果论者可能将其标准直接用于准则，而仅间接用于行为。在每个人都可以依之行动的可能准则之中，有些会是

最优的（optimific），其含义是：如果每个人都奉行这些准则，那么事情将不偏不倚地进展最好。

根据我们可称之为的

准则后果论（Maxim Consequentialism）：每个人都应当仅奉行这样的最优准则。

376 值得简短回顾一下康德的一个公式。有些康德主义者可能论证说：

（G）如果我们能够合理地意愿“每个人都奉行某一准则”为真，那么我们每一个人就都被允许奉行该准则。

（H）有些人能够合理地意愿“每个人都奉行最优准则”为真。

因此

这些人被允许奉行最优准则。

(G)是康德的自然法公式。如果(H)是正确的,那么康德的公式就允许某些人成为奉行此类最优准则的准则后果论者。

在评估这一论证的过程中,我们必须诉诸某种关于理由和合理性的观点。根据我认为我们应该接受的、那种广义的基于价值的客观观点,(H)是正确的。如果每个人都奉行最优准则,那么事情的进展既是不偏不倚地最好的,也对某些人是最好的。这些幸运者将既有不偏不倚的又有个人的理由来意愿“每个人都奉行这些准则”为真,而且其中至少一些人不会持有任何更强的、与之相冲突的理由。

有些作者主张,在运用康德的公式时,我们应当仅诉诸“避免不一致性或避免意愿中的自相矛盾”这样的理性要求。按照这一假定,(H)是正确的。会有某些人能够合理地意愿“每个人都奉行最优准则”为真,因为这绝不会包含不一致性或其意愿之中的矛盾。另有一些作者主张,理性会要求我们去意愿最好地满足我们作为理性行动者的真正需要。按照这一假定,仍会有些幸运者能够合理地意愿“每个人都奉行最优准则”为真。在这样的世界,事情将进展最好,其部分原因在于许多人作为行动者的真正需要将得到最好的满足。

按照理由的主观理论,(H)也是正确的。在这些幸运者之中,有些人会极为关心他人的福祉,并且会要求事情以最好的方式进行。如果每个人都奉行最优准则,其中有些人的欲望就会得到最好的满足。

377

合理利己主义者可能拒绝(H)。这些人认为,对我们的合理要求是选择最有利于自己的任何事。合理利己主义者可能主张,如果每个人都奉行某些并非最优的准则,那将对每一个人各自来说最好,因为其中有些行动会给这个人带来额外利益,尽管这会给他人带来更大的损害。但我认为,这一主张是不正确的。如前所示,有些幸运者极为关心他人的福祉,并且如果事情以不偏不倚地最好的方式进行,那对其中有些人将是最好的。

这也同样适用于理由与合理性的、其他任何讲得通的或广为接受的观点。按照所有诸如此类的观点,都会有某些人能够合理地意愿“每个人都奉行最优准则”为真。由此我们可以宣称,康德的原版自然法公式允许一些人成为准则后果论者。

由于此类道德主张应当适用于每个人,而康德的公式只允许一些人成为准则后果论者,因此这就是一种反驳。我们可称之为相对主义反驳(the Relativism Objection)。要答复这一反驳,我们可以修正康德的公式,以便它不是诉诸行动者能够合理地意愿什么,而是诉诸每个人能够合理地意愿什么。

这种经修正的公式具有适用于每个人的蕴含。

我已论证过，我们还有其他的强理由以种种方式修正康德的公式。这些修正将我们引向康德式契约论公式。因此，现在我们可以问这个公式蕴含着什么。

57 康德主义论证

在每个人都可能接受的原则之中，有些可能是

UA最优的（UA-optimific），其含义是：它们作为原则被普遍接受将使事情进展最好。

根据规则后果论的被普遍接受（universal acceptance）版本，或者说

UARC：每个人都应当遵循这样的最优原则。

378

在我们考察某些种类的情况时，可能存在两种或两种以上的最优原则，它们之间有着重要的区别。由此规则后果论就不得不以某种其他的方式在这些原则之间选择。这个问题最好容后再论。于是我们在此就可以假设，仅存在一组UA最优原则。

康德主义者可以论证道：

（A）每个人都应当遵循这样的原则，其被普遍接受是每个人都能够合理地意愿或选择的。

（B）每个人都能够合理地选择他们有充足理由去选择的任何东西。

（C）存在某些UA最优原则。

（D）这些是每个人都有最强的不偏不倚理由去选择的原则。

（E）任何人选择这些原则的不偏不倚理由，都不会决定性地弱于与之相冲突的任何相关理由。

因此

（F）每个人都有充足的理由去选择这些最优原则。

(G)不存在任何其他的、每个人都有充足理由去选择的、明显地 (significantly) 非最优的原则。

因此

(H) 仅有这些最优原则是每个人都有充足理由去选择的，从而能够合理地选择的。

因此

每个人都应当遵循这些原则。

这个论证有效。(A)是康德式契约论公式。因此，如果该论证的其他前提正确，那么该公式就要求每个人都遵循这些最优原则。我们可称之为规则后果论的康德主义论证 (Kantian Argument for Rule Consequentialism)。

379

在运用康德式公式时，我们问的是：如果每个人假定自己有能力去选择所有人从今往后都会接受的原则，那么他们能够合理地选择何种原则。该公式诉诸的原则，是在众多这样的设想情况下每个人都能够合理选择的。我们应该假定，在做出这些选择的过程中，每个人都知道所有的相关事实。根据这种假定，就如前提 (B) 所主张的，每个人都能够合理地选择他们有充足理由去选择的东西。

我们假定：如 (C) 的主张，存在某一组UA最优的原则。在能为每个人接受的所有原则之中，这些原则是若被普遍接受就会使事情在蕴含不偏不倚理由的意义上进展最好。如果每个人都接受这些原则，那么事情的进展方式将是每个人有最强的不偏不倚理由想要的。这是据定义为真的。因此，如前提 (D) 所主张的，这些原则是每个人都有最强的不偏不倚理由去选择普遍接受的。

根据前提 (E)，任何人选择这些原则的不偏不倚理由，都不会决定性地弱于与之相冲突的任何相关理由。这个前提需要辩护。如果我们是在从不偏不倚的视角去选择原则，那么最优原则是每个人都最有理由去选择的。但在这一康德式公式所诉诸的思想实验中，我们不是从不偏不倚的视角去选择的。我们的选择会影响自己的生活，以及那些与我们有着紧密关系的其他一些人如我们的近亲和爱人的生活。因此，我们可能有强烈的、个人的和有偏向的理由不选择

最优原则。

380

要确定是否每个人都能够合理地选择这些原则，我们必须知道可选原则是什么。在此考察其他那些明显地非最优的原则就够了，这里的“明显地非最优的原则”意指，其被普遍接受将使世界的未来走向在某些意义上要坏得多。我们没必要将最优原则与仅是略显（slightly）非最优的原则相比较，因为后者被接受只是使事情往仅仅是稍坏的方向发展。如前所示，我们应该先弄清大致轮廓。细节容后再论。

58 自利的理由

在问前提（E）是否正确过程中，我们应该考察不选择“每个人都接受最优原则”的最强理由，它们是任何人都可能持有的。在不选择这些原则的理由之中，有些理由可能产生于有关自身福祉的事实。如果每个人都接受最优原则，那对某些人会很糟糕。他们有强烈的自利理由不选择这些原则。

我或许就属于这种人。再设想：

救生船情形（Lifeboat）：我被困于一块岩石上，而有五人被困于另一岩石上。这两块岩石都将为正在上涨的潮水所淹没。在此之前，你可以用一艘救生船，要么救我，要么救那五人。对你来说，我和这五人都是陌生人，我们也互不相识，而且我们在其他相关方面也相似。我们都很年轻；如果死去，都将损失多年的幸福人生。

任何一种最优原则都会要求你去救那五人，因为死的人越多越糟糕。根据一条这样的原则，我们可称之为：

数量原则（Numbers Principle）：我们能够救两组人中的一组，他们对我来说都是陌生人，且在其他相关方面相似，此时我们应当救有更多人的那一组。

381

接着假设我的岩石靠你更近。根据

就近原则（Nearness Principle）：在这种情况下，我们应当救离

我们较近的那一组。

如果每个人都接受数量原则，而不是就近原则，那在许多其他情况之下一些人就会奉行数量原则，从而更多人的生命会被救。这一事实会给予我以强的不偏不倚理由去选择每个人都接受数量原则。但我知道，如果我做出这个选择，你就会奉行该原则，救那五人，而我将死去，由此损失多年的幸福人生。这一事实会给予我以强的自利理由去选择就近原则，因为由此你就会救我。根据前提（E），这些自利理由不会决定性地强于或胜过我选择数量原则的不偏不倚理由。这正确吗？

按照理由的主观主义理论，答案取决于我的欲望或目标。我如果对他人的福祉怀有足够的关心，那就能够合理地选择每个人都接受数量原则。但我们如果是主观主义者，那就必须拒绝康德式公式。在大多数情况下，没有任何原则是每个人都有基于欲望或基于目标的充足理由去选择的。然而，如我已论证的，我们应当拒绝主观主义，接受某种客观主义的观点，后者会诉诸基于价值的、对象给予的理由。

根据一种这样的观点，

合理利己主义：我们总是最有理由去做最有利于自己的任何事。

在该观点看来，前提（E）不正确。我不能合理地选择每个人都接受数量原则，因为这一选择对我更坏。但我认为，我们应当拒绝这种观点。

根据与之完全相反的观点，

合理的无私主义（Rational Impartialism）：我们总是最有理由去做不偏不倚地最好的任何事。

在这种观点看来，牺牲我们的生命如果能够救若干陌生人，那就是对我们的合理要求。如果这是正确的，那么像救生船情形中的情况就不会对前提（E）提出任何反驳。选择每个人都接受某种诸如数量原则之类的最优原则，就将对我的合理要求。但我认为，我们也应当拒绝这种观点。

根据

广义的基于价值的客观观点 (wide value-based objective views)：有两种可能行动，其中之一使事情按照不偏不倚地更好的某种方式进行，而另一种使事情对于我们自己或是对于与我们紧密相关的其他人会变得更好。此时，我们通常有充足的理由按照其中任一种方式行动。

按照这样的观点，理性通常会允许但不要求，我们给予自己的福祉以及与我们紧密相关者（比如近亲和爱人）的福祉以大得多的权重或强优先权。我认为，我们应当接受某些与此类似的观点。

按照在我看来最可行的观点，我们如果能够救自己的或是几个陌生人的生命，那就有充足的理由按其中任一种方式行动。在救生船情形中，我能够合理地选择你救我，但反而也能够合理地选择你救那五人。因此，我能够合理地选择每个人都接受数量原则。

383 根据某些更为利己的客观观点，给予自身福祉以强优先权是对我们的合理要求。我没有充足的理由去放弃我的生命，除非由此我会救多达上百乃至上千的他人生命。但在康德式公式所诉诸的思想实验中，我将有能力去选择“哪些原则是每个人从今往后都会接受的”。我选择的原则就会被数以亿计的人所接受。如果我选择每个人都接受数量原则而不是就近原则，那么我的选择就会影响到，人们随后在其他众多类似情况下将如何行动。尽管我将死去，但我的选择将间接地挽救至少成百万的他人。如今每年至少有成百万的人在本来可以轻易得救的情况下死去。因此，即使按照这些更为利己的观点，我也有充足的理由放弃一己之命来救大量的他人。

以上情况仅为一例。但如我所认为的，如果我能够甚至以自己的生命为代价去合理地选择这一最优原则，那么类似的主张在众多如下的情形之中应全都适用：其中由于风险较低，任何人选择最优原则都不会涉及在其自身福祉上做出如此之大的牺牲。

接着假设我的看法是错的。我曾主张，我们应当拒绝合理利己主义。但存在与此相关的、更有理由的又一种观点。按照该观点，我们通常能够合理地选择去承受某种重担，只要由此能够让众多的他人免除类似的重担。然而，在这种负担之重如英年早逝从而损失多年幸福人生的情况下，上述主张就不正确了。我不能合理地选择数量原则，因为我不能合理地选择放弃自己的生命，而无论我的选择能救多少人。我们可将这种观点称作大赌注的利己主义 (High

Stakes Egoism)。

如果该观点是正确的，那么救生船情形就不仅对规则后果论的康德主义论证之前提(E)，而且对康德式契约论公式都提出了反驳。正如我不能合理地选择“要求你救那五人而不是我”的任何原则一样，那五人也不能合理地选择“要求你救我而不是他们”的任何原则。在如此种种情况之下，将不会产生任何每个人都能够合理地选择的原则。因此，也不会产生任何康德式公式会要求我们去遵循的原则。如果我们可以救一个陌生人或是另外一百万人，该公式也会允许我们按照其中任一种方式行动。这是一个不可接受的结论。

我认为，大赌注的利己主义是不正确的。但值得说明的是：如果该观点正确，那么我们能够如何回应对康德式公式的这一反驳。

契约论者诉诸每个人都会合理地选择的原则，只要我们在作选择的时候是以某种使得我们的选择足够不偏不倚的方式进行。罗尔斯提出，要实现这样的不偏不倚，我们就应该诉诸那些每个人从某种无知之幕后都会合理地选择的原则，无知之幕后防止我们了解关于自己及其处境的特定事实。我曾主张，在运用康德式契约论公式时，我们不需要这样的无知之幕后。总会存在某种适当的原则，它是每个人即使在充分知情的情况下也都能合理地选择的。

384

我们现在的假设是，在一种情形下我的主张是错误的。在此情形下，有两组陌生人的生命待救，其中一组人多些，我们可以救其中任一组。根据大赌注的利己主义，这两组人在选择适用于该情形的原则时，对他们的合理要求将是，给予救自己的生命以绝对的优先权。康德式公式在此就失败了，因为这些人的选择是全然自利的。为免于这一反驳，我们可以修正该公式。在将之运用于这样的情形时，我们可以诉诸这些人从不偏不倚的视角来合理地选择的原则。或者我们可以部分地效仿罗尔斯，添加一个局部的(local)无知之幕后，由此这些人就不知道他们是处在人多的还是人少的那一组。按照康德式公式的这两种版本，这些人就全都能够合理选择某种最优原则，该原则会要求我们去救人多的那一组。

通过其所诉诸的原则全都这样被挑选出来——要么从不偏不倚的视角，要么从全局的(global)无知之幕后来选择——康德式公式可以得到更彻底的修正。但由此将如我在第五十二节所说明的，这样会减弱该公式的吸引力。而且做这样的修正也毫无必要。大赌注的利己主义仅适用于如下情形：如果我们选择某种最优原则，那么该选择会给我们施加极为巨大的负担，比如说英年早逝或不得不忍受持久的折磨。我们能够合理地选择去接受较弱一些的伤害，比如

说变聋或断腿，只要我们的选择将间接地使众多他人免于类似的伤害。因此，我们仍可以宣称，在人们的利益相冲突的几乎所有情况之下会存在某种原则，它是所有这些人即使充分知情且从其有所偏袒的实际视角去看都会合理选择的。

如果如我所认为的，我们应当拒绝大赌注的利己主义，那么按上述方式对康德式公式作哪怕是部分的修正也是没有必要的。

59 利他理由与道义式理由

385 在不选择最优原则的理由之中，另有一些理由可能来自有关某些他人的福祉之事实。设想

救生船情形二 (Second Lifeboat)：你可以救你的孩子，或者救五个陌生人。

我们可能认为，就你救那五个陌生人而言，你即使能够合理地放弃自己的生命，也不能合理地放弃你的孩子的生命，也不能合理地选择我们全都会接受某种要求这一行动的最优原则。由此看来，这种情况可能对前提 (E) 构成反驳。

然而，最优原则并不会要求你救那五个陌生人，而不救你的孩子。假设我们所有人都接受且奉行某种这样的原则：它要求我们不优先考虑让自己的孩子免于死亡或小些的伤害。在这样的世界，事情在某些意义上会进展更好，因为有更多的孩子得救，更少的孩子受伤害。但这些好效果极为得不偿失：如果我们都持有如此行动所需的动机，情况就会更糟糕。不优先考虑让自己的孩子免于伤害这一点要为真，我们对自己孩子的爱就必须弱得多。这种爱的减弱不仅其本身不好，而且有许多负面效果。鉴于种种诸如此类的事实，最优原则在许多情形下允许、许多情形下则要求我们给予自己孩子的福祉以强优先权 (strong priority)。

但是，上述反驳可能转移到另一种不同的情况。设想

救生船情形三 (Third Lifeboat)：这次是我能够救你的孩子，或者是救五个别的孩子。这六个孩子对我来说都是陌生人。

任一种最优原则都会要求我去救那五个别的孩子。而我们可能主张：

(I) 你不能合理地选择每个人都接受像这样的最优原则，因为你有决定性的理由转而这样选择：我会接受某种原则，它要求我去救你的孩子。

我们可以主张，你拥有此类决定性的理由，因为你有义务做出会救你孩子生命的选择。

通过诉诸我们的道德信念，我们可能还以某些其他方式争辩说，我们不能合理地选择每个人都接受某些最优原则。我们或许认为，如果每个人都接受这些原则，那么有时会导致我们或他人采取不当的行动。我们可能宣称，此类行动的不当性会给予我们以决定性理由，不选择每个人都接受这些原则。

然而，如我经常讲到的，在运用康德式公式或任何其他的契约论公式时，我们不能诉诸有关哪些行动为不当的信念。如果我们断言：

某一行动是不当的，因为我们不能全都合理地选择任何允许该行动的原则；

那么同时做出如下断言就是毫无意义的：

我们不能全都合理地选择任何这样的原则，因为该行动是不当的。

同时做出如下两个断言也同样是无意义的：

每个人都应当遵循某些原则，因为它们是每个人都能够合理地选择的唯一原则；

以及

这些是每个人都能够合理地选择的唯一原则，因为它们是每个人都应当遵循的原则。

如果我们将这些断言合并，康德式公式就会一无所成。因此，在运用这个公式时，我们必须搁置我们关于哪些行动为不当的信念。只有在后面的阶段

387 ——即在得出这个公式蕴含着什么之后，我们就要问：鉴于这些蕴含，我们是否应当接受该公式——我们才能诉诸这些信念。

既然不能诉诸我们关于你对你孩子的道义式信念，那我们还能以某种其他的方式为（I）辩护吗？我认为，我们可能最有道理的途径是诉诸你对你孩子的爱。与其设法搁置你对孩子的义务，不如调整情形更为简单。设想

救生船情形四（Fourth Lifeboat）：我可以救某个你所爱的人，或者是救另外五个人。对我来说，这六人都是陌生人。

任一种最优原则都会要求我去救另外五人。现在可能有人主张：

（J）你不能合理地选择每个人都接受某一最优原则，因为你有决定性的理由去这样选择：我接受某种别的原则，它要求我去救你所爱的人。

这一主张尽管有道理，但我认为不正确。

否认你有决定性的理由选择这一别的原则，这或许看来是荒谬的。罗密欧（Romeo）或伊索尔德（Isolde）^①能够合理地选择任凭朱丽叶（Juliet）或特里斯坦（Tristan）死去吗？在讨论一个类似的情形时，威廉斯写道：

对他人的深爱……不能体现不偏不倚的观点，而且……还有冒犯后者的危险……然而，除非如此之事存在，否则一个人的生命之中就不会有足够的意义或信念，以驱使他忠于生命本身。若要包括遵循公正体系在内的任何事有意义，则生命就必须有意义；除非生命有意义，否则它就不可能赋予公正体系以无上的重要性……

388 在此，威廉斯动人地拒绝了公正体系。而我并没有诉诸这种公正体系。如我刚讲过的，最优原则通常会允许或是要求，我们给予与我们紧密相关者的福

① 特里斯坦与伊索尔德是源自爱尔兰传说，见诸于不列颠的托马斯和法国诗人特罗亚等人作品的爱情悲歌。

祉以强优先权。而且就我们能够合理地选择每个人都接受这些原则这一主张而言，我并没有假定，给予每个人的福祉以同等的权重是对我们的合理要求。我的假定仅仅是，我们尽管被合理地允许给予自身的及某些他人的福祉以强优先性，但也被合理地允许给予陌生人的福祉以相当的权重。

如我关于救生船情形的主张所蕴含的，你爱的那个人能够合理地这样选择：每个人都接受某种最优原则。尽管这个人会由此死去，但这一选择间接救了其他大量的人。该事实给予此人以充足的理由做出这个选择。

我们所爱的人能够合理地选择为了有利于他人之故而承受某负担，此时这一事实并不意味着，我们能够合理地选择让爱人去承受这一负担。给予我们的爱人的福祉比我们自身的福祉以大得多的权重，这或许是对我们的合理要求。宁愿救五个、五十个乃至五百个陌生人而不救我们的爱人——我们或许没有这么做的充足理由。但在救生船情形四中，你会认识到，如果你选择每个人都接受某种最优原则，那你的选择就会间接挽救多得多的他人生命。我认为，你有充足的理由选择救其他大量的人。我同意，设想罗密欧或伊索尔德选择任凭朱丽叶或特里斯坦死去，这是荒谬的。你如果是罗密欧或伊索尔德，实际上就不会做出救其他大量人的选择。但我们经常看到，人们实际上未做他们有充足理由去做的事。既然你有充足的理由去选择某种最优原则，那么我认为，救生船情形四对于前提（E）或康德式公式并没有构成反驳。

接下来设想我的看法是错误的。可能有人宣称，在赌注如此之大的情况下，我们应当合理地给予爱人的福祉以绝对的优先权。如果这一点正确，那就不会产生适用于此类情况的、每个人都能够合理选择的任何原则。因此，也不会产生根据康德式公式每个人都应当遵循的任何原则。该公式就不会要求，我去救甚至成百万的陌生人而不救爱人。这又是一个不可接受的结论。这样的反驳类似于诉诸大赌注的利己主义。如前所示，可以通过添加局部的无知之幕来修正康德式公式。但我认为这种修正是不必要的。

389

60 致不当特征的反驳

按照某些基于价值的客观理论，有些事值得去做，有些目标值得去实现，这不依赖于或不仅仅依赖于它们对任何人的福祉所做的贡献。斯坎伦举出的例子是“友谊、其他有价值的个人关系，以及实现各种形式的卓越，比如在艺术或科学方面的卓越”。我们可称之为完善论（perfectionist）目标。

按照这样的观点，如果我们和他人拥有这些有价值的个人关系，以及实现别种形式的卓越，那将在蕴含不偏不倚理由的意义上是自身善的。最优原则可能要求我们去设法实现某些完善论目标，且帮助他人也做到同样的事。既然这是些有关事情如何进展最好的观点，这样的主张就不能给予我们以拒绝最优原则的理由。

然而，按照某些观点，我们或许还有些个人的和有偏向的（**partial**）完善论理由。它们并非自利的理由，因为要实现某一完善论目标，我们可能必须牺牲我们的诸多福祉。但这些理由可能冲突于我们“让事情依此完善论方式不偏不倚地更好”的理由。假设我可以挽救两种文稿之一：一种是我几近完成的一部伟大的小说孤本，另一种是五部由其他作者写作的、同样伟大的小说孤本。我可能拥有个人的完善论理由，不去选择任何最优原则——该原则会要求我挽救他人的而不是我的小说。但我认为，这样的理由不会胜过我选择该原则的不偏不倚理由。我能够合理地放弃我的小说，以挽救另五人的类似小说。如果我的看法是错误的，那我们可以再次修正康德式公式。但这几乎不会产生什么影响，因为这样的例子很罕见。

还存在另一种更重要的可能性。假设某种最优原则要求某些我们认为是不当的行动。在运用康德式公式时，我们既不能诉诸关于某些行动为不当的信念，也不能诉诸这些行动的不当性可能提供的道义式理由。但我们能够诉诸这些行动的特征，在我们看来，该特征致使这些行动成为不当的。于是我们可能宣称：

（**K**）这些致不当特征给予我们以决定性的非道义式理由，不采取这些行动且不选择每个人都接受要求这种行动的最优原则。

如果存在适用于（**K**）的某些行动，那就对规则后果论的康德主义论证之前提（**E**）提出了一个反驳，因为将存在我们缺乏充足理由去选择的最优原则。我们可称之为致不当特征的反驳（**Wrong-Making Features Objection**）。

该反驳正确地假定，在可能致使行动为不当的特征之中，有些也会给予我们决定性的非道义式理由。例如，如果某些行动会造成无谓的痛苦，这一事实将给予我们以不这么做的决定性理由。这些理由并非道义的，因为它们不是由“这些行动是不当的”这个事实所提供的。这些行动的不当性至多是给予我们不这么做的进一步的理由。但是，（**K**）不可能真正运用于这些行动，因为最优

原则不会要求我们造成无谓的痛苦。

(K) 似乎在运用于如下行动时最可能为真：该行动会产生好效果，但同时会让我们认为违反了关于某种待人方式之不当性的原则。回到

桥梁情形：你可以救五人之命，其唯一方法是使我跌倒于失控的火车前，由此致我于死。

假定我们相信这种行动是不当的，且其致不当特征在于如下事实：

(L) 你将杀死我，以用作救其他这些人的手段。

391

为了说明反驳(K)的一个版本，我们可能同时宣称如下两点：

(M) 最优原则会对我们提出这样的要求：在像桥梁情形的情况下杀死一人以用作救其他几人的手段，因为我们由此会让事情进展最好；

以及

(N) 如此行动的致不当特征将给予我们某种决定性的非道义式理由不这么做，从而不去选择任何会要求这种行动的最优原则。

(M) 并非显然正确。由于我此前和后面提及的各种理由，最优原则通常会允许甚或要求我们不去做使事情进展最好之事。但在此我们可以假定(M)是正确的。而只需问像(M)和(N)这样的断言能否同时正确就够了。

最优原则要是要求某些行动，如下就必须为真：

(O) 在从某种不偏不倚的视角来考察这些行动时，我们将最有理由想要每个人都这么做。

如果我们没有像这样的不偏不倚理由，那么每个人这么做就不会在蕴含不偏不倚理由的意义上更好，因而最优原则就不会要求这么做。在考察的情况所涉及的全是陌生人时，我们的视角是不偏不倚的。这对于几乎所有实际情况

来说都是真的，因为几乎每个人对我们来说都是陌生人。因此，我们也可以宣称，如果

(P) 最优原则要求某些行动，

如下就必定是真的：

(Q) 我们将最有理由要求几乎每个人都这么做。

392

按照我们现在考察的反驳，

(R) 这些行动之中有一些具有某些特征，这些特征会给予每个人以决定性的非道义式理由不这么做。

我认为，至少在大多数情况下，(P)、(Q) 和 (R) 不可能都为真。例如在运用于桥梁情形时，这些断言蕴含着：

(S) 你会有某种决定性的非道义式理由，不借助杀死我来救那五人；

另一方面蕴含着：

(T) 你也会最有理由要求或希望，某个陌生人到来代替你去做这样的事：杀死我来救那五个人。

按照这种观点，尽管每个人都会有决定性的非道义式理由不将杀死某人用作手段去救更多的人，但从从不偏不倚的视角看，每个人最有理由要求之事是，每个能这么做的人的确将杀死某人用作手段去救更多的人。我认为，这两种理由不可能如此直接地相对立。我们不可能有如此的不偏不倚理由，而要求每个人去做每个人都有如此决定性的、非道义式理由不去做的事。因此，(S) 和 (T) 不可能同时为真。

类似的断言适用于其他情况。在致使某些行动为不当的特征之中，大多

数给予我们非道义式理由不这么做。至少在大多数情况下，这些特征也给予我们以理由想要他人不这么做。这对于伤害他人的不当行动来说是最为清楚的，因为我们都有不偏不倚的理由想要他人不受到伤害。但类似的主张会应用到具有其他致不当特征的行动。例如，假定以欺骗或强制他人作为手段来谋利是不当的。这些行动的致不当特征可能给予每个人以决定性的非道义式理由不这么做。如果这是正确的，那么如下主张——从不偏不倚的视角看，我们会最有理由要求每个人这么做——也可以是正确的吗？我认为，我们的答案应该是否定的。如果在此情况下，欺骗与强制的性质给予每个人以决定性的非道义式理由不欺骗与强制他人，那么我们就也不可能也有像这样的不偏不倚理由：要求每个人在同样的情况下去欺骗或强制他人。那将是一种令人奇怪地前后不一的或在冲突的观点。而且如果我们没有像这样的不偏不倚理由，最优原则就不会要求这样的行动。下面我将为这些主张作进一步的辩护。

但也可能存在某种例外。设想

较小的恶情形 (Lesser Evil)：你得知，除非你杀我去救五人，否则格蕾和格林就会各自杀死另外两人去救那五个人。

在相信“以杀某人作为手段去救其他人”为不当的人之中，大多数会认为，即便它们如较小的恶情形那样是防止更多同种行动的唯一途径，这么做也是不当的。然而，即便该行动是不当的，我们仍会有不偏不倚的理由要求你这么做。尽管你杀死我以用作手段，这很糟糕；但如果格蕾和格林都按此不当方式各自杀死两人以用作手段，那显然会更糟糕。因此，如果得知你依此不当方式行动，由此防止了格蕾和格林两人的不当行动，我们就应当在某种审慎的意义上将这一事实视为好事。如果我们——如同在运用康德式公式时所必须做的那样——搁置有关哪些行动为不当的信念，那么类似的主张也适用。如果每个人都拥有此类决定性的非道义式理由不依某方式行动，那么如我已断言的，我们就不可能有不偏不倚的理由要求每个人都这么做。这会是一种前后不一的观点。但是，我们可以有不偏不倚的理由要求无人按此方式行动，除非这样的行动是防止更多同样行动的唯一途径。那就不会是一种前后不一的观点。

根据我们现在讨论的反驳：

(U) 最优原则要求我们按某些方式行动，尽管这些行动具有这

样的致不当特征：它们给予我们以决定性的非道义式理由，从而不采取这些行动且不选择每个人都接受这些原则。

如我已论证的，我们可以这样答复：

(V) 如果这些行动具有这样的致不当特征，那么最优原则就不会要求我们这么做，也许在如此行动是防止更多同样行动的唯一途径时可以例外。

如果(V)如我所认为是正确的，那么这个反驳就至多适用于只是很少的情况，比如说较小的恶情形。现在我要论证，即便是在这些情况下该反驳仍然无效。你如果预计会同意，那或许可以略过下一节。

61 决定性的非道义式理由

在桥梁情形和较小的恶情形中，你如果救那五人，就会借助杀死我去救。接着我们就可以问，是否如该反驳所宣称的，你的行动之致不当特征会给予你以决定性的非道义式理由不这么做。我们可以先重新考察隧道情形：在该情形中，如果你使一列失控的火车变向，

(W) 你会救那五人，但是以一种同时会杀死我的方式。

我们可以有道理认为，该事实给予你以强的非道义式理由不这么做。去做一件你知道会杀死一名无辜者的事，这是令人极其不快的。许多人认为，道德上仅仅允许而不是要求你改变火车的方向来救那五个人，其原因也许正在于此。但是，我认为如这些人将会同意的，杀某人的极端不快不会给予你以决定性的非道义式理由不这么做。如果道德上允许你使火车变向，那么尽管你由此会杀死我，但救多人生命这一事实将给予你以充足的理由依此行动。

类似的主张适用于桥梁情形，其中如果你使我跌倒于铁轨，

(L) 你就会杀死我以用作救那五人的手段。

通过杀死一名无辜者来救五人，这又是令人极其不快的。该行动的这一特征可能给予你以强的非道义式理由不这么做。然而，与隧道情形中一样，这一非道义式理由不可能决定性地胜过你会去救多人生命的理由。如果像许多人所认为的，桥梁情形与隧道情形有着显著的不同，那么我认为，这种不同就不可能在于：你将杀我以作为手段，因而你就有决定性的非道义式理由不这么做。该行动的这一特征或许给予你决定性的理由不这么做。但我认为，只有通过致使该行动为不当，才可能如此。这种决定性理由将不得不是道义的。如果这一点正确，那么我们正在考察的反驳就失败了。你不会有某种决定性的非道义式理由不这么做。

类似的意见适用于别种情况。我提议：

(X) 如果最优原则要求某些我们认为是不当的行动，那么按照我们的观点，使该行动为不当的特征或事实不会给予我们决定性的非道义式理由不这么做。真实的情况可能只是：由于使该行动为不当，这些事实会给予我们以决定性的道义式理由不这么做。

最优原则会要求多种许多人认为是不当的行动。比如说，这些原则或许要求我们一些人使用人工避孕、做或接受堕胎手术、帮助某人安乐死，或者是劫富济贫。如果我们有决定性的理由不这么做，那么我认为，这些理由将不得不是由这些行动的不当性提供的。

我们应该预料到(X)是正确的。如果最优原则要求某种行动，那我们就必定拥有强的不偏不倚理由去要求每个人都这么做。如果我们没有这样的理由，那么在每个人都这么做的事情就不会变得更好。因此，最优原则就不会要求这样的行动。我们既然拥有强的不偏不倚理由去要求每个人都这么做，就应该预料到，这些理由不可能被决定性地胜过，除非是通过如此行动为不当的事实。我将在附录C对(X)作进一步的辩护。

尽管我强烈地倾向于认为(X)是正确的，但值得再次姑且假设我弄错了。假设最优原则要求某些我们认为是不当的行动，又假设按照我们的观点，致使这些行动为不当的特征会给予我们以决定性的非道义式理由不这么做。这些信念本身不会对前提(E)构成反驳。这种反驳必须宣称：

(Y) 这些致不当特征同时也会给予我们以决定性的非道义式理

由，不去选择每个人都接受要求这些行动的最优原则。

(Y) 蕴含着，存在某种最优原则是我们不会有充足理由去选择的，由此只有(Y)才会不利于(E)。

(Y) 作为一种主张，并非关于按某些方式行动的理由，而是关于选择每个人都接受某原则的理由。这是相当不同的问题。例如，想想某类这样的行动：它对我们来说是坏的，但会给他人带来更大的利益。我们即使拥有强理由不这么做，但可能有决定性的理由既要求每个人都这么做，又选择每个人都接受某种要求如此行动的原则。譬如说，如果每个人都这么做，那对于包括我们在内的每个人都可能更好。

在运用于那些违反道义论限制 (deontological constraint) 的行动时，(Y) 似乎最有可能正确。我们的主要例子是桥梁情形。我们同时做出两点假设：其一是在此情况下，最优原则会要求你救那五人；其二是由于你将杀死我以作为手段这一事实，这将使得该行动成为不当的。根据(Y)，这个事实会给予你某种决定性的非道义式理由，而不去选择每个人都接受任何诸如此类的最优原则。我们应该问的是，这种理由可能是怎样的。

这种理由既然必须是非道义的，那就不能由如此行动的不当性来提供。我们可以再次诉诸通过杀死一名无辜者来救多人的可恶性 (awfulness)。我们可以有道理认为，这种行动的可恶性给予你以某种强的非道义式理由，由此你希望在道德上不被要求这么做。但如我们所见，在像隧道情形的情况下，这种理由并非决定性的，因为你会有充足的理由以同时杀死我的方式去救那五人。

397

而如果在桥梁情形中，最优原则要求你用杀我的方式去救那五人，那就必须是因为相关事实给予你以不偏不倚的理由，要求每个人在如此情况下都这么做。这些事实也会给予你以理由，想要每个人都接受某种要求他们这么做的原则。我认为，这些不偏不倚的理由不可能决定性地弱于你个人的非道义式理由，后者想要你自己不被要求这么做。

在为上述看法辩护的过程中，我要做出某种更宽泛的 (wider)、适用于一切情况的主张。如果最优原则要求我们依某种方式行动，相关事实就必须给予我们以不偏不倚的理由，想要每个人在相关方面类似的情况下都这么做。仅在此时，每个人如果都这么做才会更好。我们既然是从不偏不倚的视角来考察几乎所有有这样的情况，那就最有理由要求几乎每个人都这么做。如果我们选择每个人都接受那种要求这么做的原则，那么我们的选择就间接地导致，大多数人愿意做

我们最有理由要求几乎每个人都做的事。这些事实给予我们以强的不偏不倚的理由，去选择每个人都接受该原则。根据前提（E），这些理由不会决定性地弱于任何相冲突的相关理由。现在我们要问，是否如（Y）所宣称的，在某些情况下（E）是错误的。

在此回顾一下对（E）提出最强反驳的其他情况，将不无裨益。如果我们选择每个人都接受某种最优原则，那么这一选择要么对于我们自己要么对于我们紧密相关者（比如我们的爱人）来说会很糟糕。譬如在救生船情形中，如果我选择每个人都接受数量原则，那你就会救那五人而不救我，由此我会损失多年的幸福人生。这一事实将给予我很强的个人理由不选择数量原则。但我认为，该理由不会是决定性的。通过选择每个人都接受最优原则，我将间接挽救极大量的他人生命，从而这个事实会给予我充足的理由去做出这样的选择。

我们现在考察的是不同类型的理由。在（Y）可能适用的情况之中，相关事实将给予我们强的不偏不倚理由，既想要每个人都按某种方式行动，又选择每个人都接受要求这么做的最优原则。但（Y）宣称，这些不偏不倚的理由会被某些与之相冲突的非道义式理由所决定性地胜过。像这样的任何理由将必须比我刚才提到的个人理由——譬如我们这样的理由：不想英年早逝从而损失多年的幸福人生——要强得多。这种理由仅当强得多时才可能决定性地胜过与之相冲突的不偏不倚理由。我认为，只有第三类理由，才可能明显强于且决定性地胜过上述的这两种理由，即如此之强的个人理由和不偏不倚理由。如果我们会有一种决定性的理由不去做某选择，尽管事实是该选择要么（1）对于我们自己或爱人会好得多，要么（2）会使事情不偏不倚地得到极大的改善，那么这种理由将不得不是由如下的事实提供的：该选择在道德上是不当的。我们不可能有决定性的非道义式理由不去做该选择。如果情况就是如此，那就如我所认为的，（Y）不可能正确。因此，对（E）的这一反驳就失败了。

398

62 每个人都能够合理地意愿什么

根据前提（E），任何人拥有的、选择最优原则的不偏不倚理由都不会决定性地弱于任何与之相冲突的相关理由。在为（E）辩护的过程中，我已诉诸多种我认为正确的主张，而后做出了这样的论证：即使我错了，（E）仍然会是正确的，或者可以通过对康德式公式作某种可接受的修正而使之正确。在此意义上，前提（E）是牢靠的。

值得姑且假设我又弄错了。假设在某些情况下（Y）是正确的，因为我们会有一种决定性的非道义式理由不选择每个人都接受最优原则。还假设该反驳不可能通过对康德式公式的任何类似修正来应对。在这样的情况下，（E）将是不当的。康德主义论证无法说明，康德式公式总是要求我们遵循最优原则。我们将不得不修正该论证的结论。

399 届时该论证将在另一种意义上仍是牢靠的，因为这种修正会很细微。由于前面给出的理由，（Y）正确的情况即使存在，也会很罕见。（Y）正确仅可能出现在于像较小的恶情形那样的情况，其中最优原则所要求的行动是作为防止更多同样行动的唯一途径。既然这样的情况很罕见，康德主义论证将说明，在几乎所有的实际情况下，康德式公式都要求我们遵循最优原则。由此康德式契约论就其蕴含而言将非常接近于规则后果论。它们之间的分歧，或许少于规则后果论的某些不同版本之间的分歧。

接下来回顾，在假设（Y）有时正确时，我假设我先前的若干主张是错误的。我认为，（Y）绝不会正确。若如此，则该论证的结论就没有必要修正。

我相信，再没有别的对于（E）的强有力反驳。若如此，则我们应当接受前提（B）至（E）。每个人都会有强的不偏不倚理由去选择最优原则，并且这些理由不会决定性地弱于与之相冲突的任何相关理由。

我们既然应当接受这些主张，就应当接受该论证的第一个结论。如（F）所主张的，每个人都将有充足的理由去这样选择：每个人都接受最优原则。

根据该论证余下的前提：

（G）不存在任何其他的、明显地非最优的原则，其被普遍接受是每个人都有充足理由去选择的。

400 与前提（E）相比，这个前提要容易辩护得多。如果每个人都接受任何像这样的其他原则，事情的进展将在不偏不倚的意义上要差得多。这是这一断言——这样的其他原则明显地非最优——所蕴含的。这些事实将给予每个人以强的不偏不倚理由不去选择每个人都接受这样的原则。大多数人既然不会有与之相冲突的个人理由，那就不可能合理地做出上述选择。而且在几乎所有这样的情况下，如果每个人都接受这样的非最优原则，事情对于某些不幸者来说就还要坏得多。甚至更为清楚的是，这些不幸者不可能合理地选择每个人都接受这种原则，因为他们将既有强的不偏不倚理由，又有强的个人理由不做出这种选

择。例如在地震情形中，怀特不可能合理地这样选择：我们全都接受某种要求我保全格蕾的腿而不救怀特的命的非最优原则。而在救生船情形中，那五人之中无人能够合理地这样选择：我们全都接受某种要求你去救我而不救所有那五人的非最优原则。因此，如（G）所主张的，不存在任何其他的、明显地非最优的原则是每个人都有充足理由去选择的。

（B）、（F）和（G）合起来说就意味着：

（H）只有最优原则之被普遍接受，是每个人都有充足理由去选择，从而能够合理地选择的。

一旦与（H）合并，康德式公式就意味着，每个人都应当遵循这样的原则。我在一个注释中为这些主张做了进一步的辩护。

现在我们可以更简洁地重述这一论证。康德主义者可能宣称：

（A）每个人都应当遵循这样的原则，其被普遍接受是每个人都能够合理地意愿或选择的。

（C）存在某些原则，其被普遍接受将使得事情进展最好。

（F）每个人都能够合理地意愿每个人都接受这些原则。

（H）这些是仅存的原则，其被普遍接受是每个人都能够合理地意愿的。

因此

UARC：这些是每个人都应当遵循的原则。

（A）是康德式契约论公式，而UARC是规则后果论的一个版本。我们做出（C）的假定。我认为，我已经成功地辩护了（F）和（H）。因此，这一康德式公式要求每个人都遵循规则后果论的原则。

401

我们或许怀疑，这个论证必须有至少一个后果论前提。如果这一点为真，那么该论证就将是无趣的。我们会预料后果论的前提蕴含着后果论的结论。而且这样的论证不会为非后果论者改变他们的观点提供任何理由。

然而，该论证的前提并非后果论的。该论证假定，后果在蕴含不偏不倚理

由的意义上可能是更好或更坏的。而非后果论者可以接受这个假定。比如说，许多非后果论者认为，如果更多的人受苦或英年早逝，那么情况会更坏。这些非后果论者之所以拒绝后果论，并不是因为否认后果可能在此意义上更好或更坏，而是因为认为行动的正当性不是仅取决于有关事情如何进展会最好的事实。该论证还假定，存在某些其被普遍接受将使得事情进展最好的原则。但这个假定也不是后果论的。我们可能相信存在像这样的最优原则，但同时也相信我们应当拒绝其中的某些原则，因为它们要求或允许某些不当的行动。

既然该论证没有假定后果论为真的任何前提，那就值得解释，该论证如何有效地蕴含着其后果论的结论。

后果论者诉诸的主张是有关如何进展在蕴含不偏不倚理由的意义上是最好的。这些主张是关于从不偏不倚的视角看每个人都最有理由想要或选择什么。对后果论的最有力的反驳，来自某些我们关于哪些行动为不当的直觉信念。

契约论者诉诸这样的原则：如果我们所有人都以某种会使得我们的选择足够不偏不倚的方式在做选择，那么对于每个人来说，选择该原则就是合理的。

402 有些契约论者宣称，要实现这样的不偏不倚，诉诸这样的原则就足够：如果每个人都需要就这些原则达成协议，那么对于每个人来说，选择该原则就是合理的。另一些契约论者比如罗尔斯则添加了无知之幕。康德式契约论者实现不偏不倚的途径是诉诸：如果每一个人都各自认为有能力去选择每个人都会接受的原则，那么每个人都能够合理地选择什么。在此，不偏不倚的实现是通过一致同意的要求，而无需达成任何协议或无知之幕。在论证存在每个人都能够合理地选择的原则的过程中，我已诉诸契约论的另一特征。在运用任何契约论公式时，我们不能诉诸我们关于哪些行动为不当的直觉信念。

现在我们可以解释，上述论证如何在无需任何后果论前提的情况之下有效地蕴含着其后果论的结论。如我刚才所言：

后果论者诉诸的相关主张是：从不偏不倚的视角看，对于每个人来说选择什么将是合理的。对后果论最有力的反驳，来自某些我们关于哪些行动为不当的直觉信念。

契约论诉诸的相关主张是：按照某种使得选择为不偏不倚的方式，对于每个人来说选择什么将是合理的。在契约论的道德推理过程中，我们不能诉诸我们关于哪些行动为不当的直觉信念。

既然两种理论都诉诸对于每个人来说不偏不倚地选择什么将是合理的，而且契约论者要求我们搁置非后果论的道德直觉，那么我们应该预料到，含有某种契约论前提的有效论证可能得出某种后果论的结论。

我们可以得出又一结论。我已经宣称，存在某些对于康德的普遍法则公式的、决定性的反驳。为了免于这些反驳，康德的公式必须被修正。就其经修正的最佳形式而言，该公式要求我们遵循这样的原则：其被普遍接受是每个人都能够合理地意愿或选择的。我已经论证，绝不存在每个人都能够合理地选择的、明显地非最优的原则。因此，除非如我也论证过的，每个人都能够合理地选择最优原则这一点为真，否则该公式就不可能成功。除非在其经修正的形式中，康德的普遍法则公式蕴含规则后果论，否则该公式就不可能成功。

403

第十七章 结论

63 康德式后果论 (Kantian Consequentialism)

404 接着回到行动后果论，或

AC: 每个人都应当总是做或努力做使得事情进展最好的任何事。

这是UA最优——即，其被普遍接受将使事情进展最好——的原则吗？

如西季威克之所论，答案是否定的。如果每个人总是都努力做使事情进展最好的任何事，这些努力通常就会失败。在预测可能行动的效果时，人们经常会犯错，或者是以自利的方式欺骗自己。例如，我们会轻易相信，我们的偷窃或撒谎是有正当理由的，因为我们误以为，我们的获利将超过我们的行动给他人带来的负担。如果我们全都是行动后果论者，那就还会损害或削弱某些有价值的实践或制度，比如说需要信任的许诺实践。如果每个人都持有行动后果论者的动机，那在其他方面会很糟糕。每个人几乎总是努力使事情进展最好，这一点要为真，我们大多数人就不得不失去太多的深爱、忠诚、个人目标，以及其他动机，这些动机是我们诸多幸福之所系，且同时在其他方面赋予我们的人生以价值。由于如此种种理由，我们可以宣称：

405 (A) 如果每个人都接受AC，那么与如果每个人都接受某些其他的原则相比，事情会变得坏些。

这些其他的UA最优原则，将与常识道德的原则部分地重叠。例如，这些

原则通常要求我们不去偷窃、撒谎或违背诺言，即便可以预期这样的行动会使事情进展最好也如此。这些原则允许我们给予自身的福祉以某种强的优先权。而且它们经常允许且不时会要求我们，给予某些他人——比如，我们的近亲与朋友，以及诸如我们的学生、病人、客户、同事、邻居以及为我们所代表者之类的以其他方式与我们可能相关者——以某种强的优先权。既然AC并非其被普遍接受会使事情进展最好的原则，康德式公式就不要求我们成为行动后果论者。

我们一直在讨论规则后果论的普遍接受版，或者说UARC。根据该理论的一个不同版本，我们可称之为

UFRC: 每个人都应当遵循如下情况为真的原则：如果这些原则被普遍遵循，那么事情会进展最好。

我们可将这样的原则称为UF最优的。当做到了某原则要求的事情时，我们就遵循了该原则。例如，如果我们总是做会使事情进展最好的任何事，那就在遵循AC。

我们也一直在讨论可称之为康德式契约论的接受版（**acceptance version of Kantian Contractualism**），或者说AKC。根据康德式公式的一种不同版本，我们可称之为

FKC: 每个人都应当遵循这样的原则，其被普遍遵循是每个人都能够合理地意愿或选择的。

前面讨论的康德主义论证可被修正，以说明：

(B) 仅有UF最优原则之被普遍遵循是每个人都能够合理地意愿的。

406

康德式公式的这一另版由此要求我们遵循这样的原则。

根据某些作者的观点，行动后果论的原则是UF最优的。例如，卡根（Kagan）宣称：

(C) 如果每个人通过做会使事情进展最好的任何事而总是遵循

AC，那么事情会进展最好。

这一断言或许看上去无可争辩。而如果该断言正确，康德式公式的这一版本将要求我们成为行动后果论者。

我认为，(C) 不正确。在问如果每个人都遵循AC，事情是否会进展最好时，我们应该考虑这样的世界将区别于其他可能世界——其中每个人都遵循各种其他原则——的所有方式。我们应该纳入考虑的，不仅是人们的行动效果，而且包括人们依此方式行动的意图与动机所产生的效果。出于西季威克所给出的某些理由，我们可以宣称：

(D) 如果每个人都总是遵循AC，那么与如果每个人都总是遵循某些其他的原则相比，事情会变得坏些。

如果每个人总是做使事情进展最好的任何事，那么每个人的行动在大多数情况下会有最好的可能后果。与如此情况——每个人总是努力去做使事情进展最好的任何事，但这样的努力经常失败——相比，事情的进展会好些。但我认为，每个人行动的好效果又会被这样的可能后果——如果我们所有人都持有导致我们遵循AC的动机，情况就会变得更坏——所抵消。如前所示，就失去诸多深爱、忠诚及个人目标而言，我们许多人将失去太多使生活有价值的东西。因此，康德式公式的这一版本不会要求我们成为行动后果论者。

然而，该公式的确要求我们遵循UF最优的原则。而且与UA最优原则相比，这样的原则更类似于AC。因此，康德式公式的这一版本所支持的道德观点，显然更接近于行动后果论。

407

为了将康德式公式的两个版本都涵盖进来，我们可以将康德式契约论重述为

KC：每个人都应当遵循这样的原则：每个人都能够合理地意愿该原则成为普遍法则。

要么由于被普遍接受，要么由于被普遍遵循，原则可以成为普遍法则。

既然KC和RC的这些不同版本具有不同的蕴含，我们或许就不得不在它们之间做出选择。要做出这一选择，我们将不得不考虑在此暂不考察的若干问题。但我要提到某种可能性。我已经讲过，我们应当区分不同意义的“应当”

(ought) 与“不当的”(wrong), 我们可以在道德理论的不同部分去使用它们以回答不同的问题。做出类似的其他区分是值得的。例如, 如果我们假定都会成功, 那么我们在理想情况下全都应当做什么; 这是一个问题。我们对这个问题的回答将是我们的理想行动理论(ideal act theory), 或者是有些人所称的完全服从理论(full compliance theory)。当得知其他某些人会行动不当时, 我们应当做什么; 这就是另一个问题。有些人称之为我们的局部服从理论(partial compliance theory)。我们还可以问, 在将各种其他事实——比如说, 有关他人可能犯错的事实, 以及有关人们的动机、欲望和性情的事实——纳入考虑时, 我们应当努力做什么。又一个问题是, 我们应当拥有哪些动机, 以及应当倾向于做什么。我们对这个问题的回答将是我们的动机理论(motive theory), 其本身又分为理想的与非理想的部分。我们如果是康德式契约论者和规则后果论者, 可能就没必要在KC和RC的不同版本(至少是某些版本)之间做出选择, 因为我们可以在道德理论的不同部分诉诸这些不同的版本, 并使用不同意义的“应当”和“不当的”。

可能还存在别的复杂性。我已经假定, 存在一组UA最优的原则, 且存在另一组UF最优的原则。如果存在两组及两组以上这样的原则, 各组之间有着显著的区别, 那么我们就不得不以某种其他的方式在这些原则组之间做出选择。还有着其他的可能性, 但在此暂不考虑。

408

现在我们可以回到康德观点的另一部分。根据我所称的康德的

至善公式(Formula of the Greatest Good): 每个人都应当力求促进一种由普遍的德性和应得的幸福(deserved happiness)组成的世界。

康德宣称, 通过遵循如康德的其他公式所描述的道德法则, 我们能够最大限度地促进这种世界。我已论证, 这些公式之中有一些在康德式契约论之中得到了最佳的修正和合并。因此, 康德原本可以宣称:

KC: 每个人都应当遵循这样的原则: 每个人都能够合理地意愿该原则成为普遍法则。

(E) 每个人都能够合理地意愿如下的原则成为普遍法则: 该原则之成为普遍法则将使世界最接近其理想状态, 由此使得事情进展最好。

(F) 该理想状态将是一种由普遍的德性和应得的幸福组成的世界。

因此

每个人都应当遵循这样的原则：其成为普遍法则将最大限度地促进这样的理想世界。

这一论证将给予康德的道德理论以最为和谐一致的形式。康德的至善公式将描述每个人都应当力求实现的、单一的终极目的或目标，而康德式契约论则描述其被普遍接受或遵循将最能实现该目标的道德法则。

409 在该论证的前提之中，KC是康德式契约论。第十六章中的康德主义论证经过某些修正，可以转变为对(E)的辩护。(F)是康德对其所谓至善的理想世界的描述。

我已论证，我们应当修正(F)。康德宣称，如果人们拥有与其应得相比更多的幸福或较少的痛苦，那就是不好的。但康德也宣称：

(G) 如果我们的所有决定纯粹是时间中的事件，那就无人应得痛苦。

我已论证，我们应当接受这一主张。如我所言，我们可以补充说：

(H) 我们的所有决定纯粹是这样的事件。

因此

(I) 无人应得痛苦。

也没有任何人应得更少的幸福。如果我们略去康德关于应得的主张，那么康德的理想世界就是由普遍的德性和幸福组成的。在考察并非理想世界的过程中，我们必须再做决定，哪些世界更接近于理想。我认为，不仅痛苦较少而幸福更多总是更好的，而且更多的幸福降临于幸福较少或受苦较多的人也总是更好的。我们可能补充说：我们的福祉不仅仅在于幸福和避苦，而且事情进展如

何，部分地依赖于其他并非与任何人的福祉相关的事实。

康德关于理想世界的主张提出了另一个问题。要问如何能够最接近康德的理想，我们就必须比较德性之善与幸福。在一种观点看来，德性之善更大，其大无穷；乃至任何人的德性稍增或邪恶稍减，都胜过得到任何数量的幸福或避免任何数量的痛苦。我认为，这种观点要看上去有道理，我们就必须假定：我们拥有某种自由，这种自由能够让我们以某种蕴含应得的方式为自己的行为负责。如我已主张的，如果不可能存在这样的自由，那我们就应当接受一种极为不同的观点。例如，如果某人由于是名残忍的谋杀犯而是道德上的坏人，那么这对于该谋杀犯、其受害者及其他人而言就都是坏的，而且还会是一种坏的事态，是我们全都有理由感到遗憾和设法防止的事态。但我相信，某人是一名残忍的谋杀犯，其坏处（badness）在相关方面类似于某人精神失常的坏处。这样的坏处能够轻易地为巨大的痛苦之坏处所胜过。

对应得的这种拒绝或许看上去让我们远离康德的观点。但康德有时做出了此类主张，比如当他提到

最高的目的是全人类的幸福。

以及在一个早期讲座中，康德说：

如果我们按这样一种方式为人处事：如果其他每个人都这样为人处事，就会产生至大的幸福，那么我们就是在为配享幸福而如此地为人处事。

康德在此坚持一种规则后果论的享乐主义版本。

现在我要总结这些论断。要么由于被普遍接受，要么由于被普遍遵循，道德原则可以成为普遍法则。我已讲过，康德主义者可以论证道：

KC：每个人都应当遵循这样的原则：每个人都能够合理地意愿该原则成为普遍法则。

(J) 存在某些原则，其成为普遍法则将使事情进展最好。

(K) 这些是每个人都能够合理地意愿其成为普遍法则的仅有原则。

因此

RC: 每个人都应当遵循这些最优的原则。

411 KC与RC都是对康德式契约论与规则后果论的最一般陈述。我们在假定(J)为真。我相信已成功地为(K)做了辩护。因此, 康德式契约论蕴含着规则后果论。

既然这是正确的, 这些理论就可以合并。根据我们可称为的

康德式规则后果论(Kantian Rule Consequentialism): 每个人都应当遵循最优原则, 因为它们是每个人都能够合理地意愿其为普遍法则的仅有原则。

64 攀登山峰

接下来回顾, 根据

斯坎伦的公式: 每个人都应当遵循无人能够合情理地拒绝的原则。

康德主义者可能论证道:

(A) 如果我们不能合理地意愿两种原则之一种成为普遍法则, 那就必定存在给予我们以强烈反驳该原则的事实。

(B) 如果每个人都能够合理地意愿另一原则成为普遍法则, 那就无人能对该替代原则提出同样强烈的反驳。

(C) 既然我们对第一种原则的反驳要强于任何人对该替代原则的反驳, 那么我们就能够合情理地拒绝前者。

(D) 一旦仅有一种相关原则是每个人都能够合理地意愿其为普遍法则的, 那么任何人对该原则的反驳都不可能强于对每种替代原则的最强反驳。

412 (E) 无人能够合情理地拒绝某一原则, 只要对每一种替代它的原则存在更强的反驳。

因此

(F) 一旦仅有一种相关原则是每个人都能够合理地意愿其为普遍法则的原则，那就无人能够合情理地拒绝该原则。

(G) 既然对每一种替代原则都存在更强的反驳，那么这些替代原则都可以被合情理地拒绝。

因此

(H) 一旦仅有一种相关原则是每个人都能够合理地意愿其为普遍法则的原则，那它就是无人能够合情理地拒绝的、仅有的相关原则。

(I) 仅有一组原则是每个人都能够合理地意愿其为普遍法则的原则。

因此

它们是无人能够合情理地拒绝的仅有原则。

我们可将此称为趋同论证 (Convergence Argument)。如果该论证是合理的，那么康德式契约论与斯坎伦式契约论就可以合并。无人能够合情理地拒绝的原则，就同于每个人都能够合理地意愿其为普遍法则的原则。

该论证的适用对象，并非斯坎伦目前的理论，而是我所认为的、斯坎伦式契约论的最佳版本。在第二十一至二十三章中，我对这一看法做了辩护，并进一步讨论了该论证。

如我已经论证的，这一合并的理论也可以包含规则后果论。根据我们可称为的

三重理论 (Triple Theory)：某一行动是不当的，当且仅当或恰好当 (just when) 这样的行动为某原则所不容，该原则是

(1) 这样的原则之一：其成为普遍法则将使事情进展最好；

(2) 这样的仅有原则之一：其成为普遍法则是每个人都能够合理地意愿的；

以及

(3) 一种无人能够合情理地拒绝的原则。

更简洁地说，

TT：某一行动是不当的，恰好当这样的行动为某原则所不容，该原则是某种最优的、唯一地普遍可意愿的，以及不可合情理地予以拒绝的原则。

我们可将此称作受三重支持的原则 (**triple supported principles**)。如果某一原则拥有这三种特性中的任一种而没有另两种，我们就不得不问，这些特性之中哪一种最具有道德重要性。但我已经论证，这三种特性为所有且只有同样的原则所拥有。如果这一点为真，那我们就可以宣称：

(**J**) 道德原则不可合情理地予以拒绝，恰好当它们是唯一地普遍可意愿的；以及他们是唯一地普遍可意愿的，恰好当它们是最优的。

我们还可以宣称：

(**K**) 一旦某原则是最优的，那就使得它成为普遍可意愿的仅有原则之一；

以及

(**L**) 一旦某原则是普遍可意愿的仅有原则之一，那就使得它成为无人能够合情理地拒绝的原则之一。

我们可以补充说：

(**M**) 一旦某些行动为某原则所不容，该原则是某种最优的、普遍可意愿的和不可合情理地拒绝的原则，那就使得这些行动在他人面

前无法辩护。

(N) 这样的行动是该受指责的，且会让其行动者有理由感到悔恨，让他人有理由表达义愤。

(O) 每个人都有理由绝不按这样的方式行动。这些理由总是充足的且经常是决定性的。

出于我先前给出的理由，这种三重理论应该宣称其所描述的不是不当性本身，而是使得行动为不当的某种特性或事实。存在多种其他的、更具体的致不当特性或事实，比如造成无谓的痛苦或为一己之便而强制他人的特性。三重理论应该宣称其所描述的是单一的、高阶的致不当特性，它可以纳入或汇聚所有其他这样的特性。这种高阶的特性是一种为某原则所不容的、复杂的性质，对于该原则而言，(1)、(2)、(3) 为真。当行动具有其他某些属性时，它们就不容于受三重支持的原则，并且我们就可以主张所有这些事实都使得这些行动成为不当的。我们可以补充说，这些事实中的每一种都给予每个人进一步的理由不这么做。

如果接受这种三重理论，我们应该承认，就解释许多种行动为什么是不当的而言，我们没必要宣称这些行动不容于受三重支持的原则。在某些情况下，这样的宣称不仅没必要，而且令人困惑和不安。它类似于这样的情形：在某强奸案或谋杀案之后，我们不应当问“如果每个人都这么做会怎样？”，或“如果每个人都相信这样的行动是可允许的会怎样？”有些行动所经受的反对，比康德的公式或任何版本的契约论或规则后果论所提供的反对都既更清晰也更强烈。

然而，在许多其他情况下，某一行动是否被允许或不容于受三重支持的原则，这样的提问可能是有益的。例如，如下行动是否为不当可能是不清楚的：违反某一法律、为实现某种善目的而撒谎、以某种方式强制某人以促进这个人或他人的善、偷窃某个其物主从不使用的东西、未能帮助某个极需帮助的人、放弃投票，或者是在一个人口过剩的世界，生养两个以上的孩子。如果任何此类行动将为某原则所不容——该原则是其被接受将使事情进展最好的原则之一、其成为普遍法则则是每个人都能够合理地意愿的仅有原则之一、或一种无人能够合情理地拒绝的原则，那么这样的事实就会对该行动提出某种最强烈的反对。

接下来回顾，按照三重理论，某一行动是不当的，恰好当这样的行动不容于受三重支持的原则。存在若干低阶的致不当特性以及不容许拥有这些特性的行动之原则。三重理论做出关于所有这些特性和原则之共性的断言。如果该理

论的断言为真，那将对如下问题提出更深层次的解释：为什么这些原则是正当的，以及为什么这些行动是不当的。如斯坎伦所述，这样一种理论的一个目的在于，提供“关于不当性的一般标准，将这些更具体的致不当特性联系起来并予以解释”。

道德理论要取得成功，就必须具备讲得通的蕴含。三重理论具有许多这样的蕴含。但在得出该理论蕴含什么且对一切相关事实和论证深思熟虑之后，该理论有可能冲突于我们关于某些行动不当性的直觉信念。如果存在许多这样的冲突，或者是这些直觉信念很强，那我们就可以有拒绝这一理论的正当理由。如果转而是这些冲突显然不那么深刻、不那么常见，那我们就可以修正我们的某些直觉的道德信念，而可证成地遵循该理论。

我们不仅拥有关于哪些行动为不当的直觉信念，而且拥有关于哪些原则或理论可能是正确的直觉信念。因此，任何成功的原则或理论不仅要具备讲得通的蕴含，而且必须其本身是可行的。只有这样的原则或理论，才能支持我们更为具体的道德信念。

康德式契约论通过了这种检验。如果某一行动不容于其成为普遍法则是每个人都能够合理地意愿的仅有原则之一，那就可以有理由宣称这一事实就是使得该行动为不当的事实之一。

416 斯坎伦式契约论可能看上去不仅是有理由的，而且是无可争辩的。设想我宣称：

尽管我的行动为某种无人能够合情理地拒绝的原则所不容，但我否认这样的行动是不当的。

这一断言可能看上去近乎自相矛盾。我尽管在拒绝该原则，但似乎又承认，这种拒绝是不合情理的。而如果我对该原则的拒绝是不合情理的，那么这种拒绝就不可能得到辩护。因此，我无法可证成地否认，这样的行动是不当的。而如果斯坎伦的公式看来无可争辩，那是因为该公式没有明确地纳入道义式信念的限制。该公式更完整的陈述可能是：

某一行动是不当的，恰好当这样的行动为某种无人能够合情理地拒绝的原则所不容；但其根据不同于（other than）这样的信念：该原则是错误的，因为它不容许那些并非不当的行动。

如下的断言不会陷入自相矛盾：某种行动并非不当的，因此，尽管该行动为如此这般的原则所不容，但该原则是错误的。

我相信，康德式契约论可以与最优版本的斯坎伦式契约论合并。但我对于这一信念的论证可能失败。由此我们将不得不在这两种理论之间做出选择。

然而，康德式契约论仍可以与规则后果论合并。我已经论证

(K) 一旦某一原则是最优的，那就使得它成为这样的原则之一：每个人都能够合理地意愿该原则成为普遍法则；

以及

(P) 绝不存在每个人都能够合理地意愿其成为普遍法则的其他原则。

如果上述主张是正确的，那么康德式契约论与规则后果论就若合符节了。

417

在三重理论的组成之中，规则后果论在一定意义上最难得到辩护。有些规则后果论者诉诸如下主张：

(Q) 最终要紧的仅在于，事情进展如何。

这一主张本身很有道理，也没有受到我所给出的任何论证的挑战。如果我们拒绝(Q)，那是因为该主张支持行动后果论，而行动后果论与我们关于哪些行动为不当的直觉信念之间的冲突过于频繁或过于强烈。规则后果论与这些直觉信念之间的冲突则要少得多。但如果规则后果论诉诸(Q)，那么其观点将面临某种强有力的反对。按照反对的观点，尽管最佳原则是最优的原则，但正当的行为不是最优的行为，而是为最佳原则所要求或允许的行为。按这样的方式行动是不当的：尽管我们知道这样的行动会使事情进展最好，但它为最优原则所不容。我们可以有理由反对如下观点：如果最终要紧的仅在于事情进展如何，那么做我们知道会使事情进展最好之事就不可能是不当的。

规则后果论反而可以建立在康德式契约论的基础之上。在此作为基本的东西，并非关于“最终要紧的是什么”的信念，而在于如下信念：我们应当遵循其被普遍接受或遵循是每个人都能够合理地意愿的原则。康德式规则后果论

没有假定最终要紧的仅在于事情进展如何，因而其观点避免了我刚才所述的反对。这些人认为，一旦行动为不当，那就不仅仅或不主要是由于这样的行动不容于最优原则之一；这些行动之为不当，还由于它们不容于这样的仅有原则系列之一：每个人都能够合理地意愿该原则系列成为普遍法则。

418 如果像我所主张的那样，康德式契约论蕴含着规则后果论，那并不会导致由此得出的观点就完全是后果论的。尽管就其关于我们应当遵循哪些原则的主张而言，这种观点是后果论的，但无论是其关于我们为什么应当遵循这些原则的主张，还是其关于哪些行动为不当的主张，它都不是后果论的。我们或许可以说，这种观点仅仅是三分之一的后果论。

我在本卷已经论证，随着某些修正和补充，康德最重要的主张如下：

(R) 每个人都应当仅以人们能够合理地同意的方式来对待每个人。

(S) 每个人都应当带着尊重的态度去看待每个人，而绝不将他们仅仅视为手段。甚至是道德上最恶的人，也拥有与其他任何人同等的尊严或价值。

(T) 如果我们的一切决定都纯粹是时间中的事件，我们为自身行动负责的方式就不可能是使我们应受苦或少享福。

(U) 每个人都应当遵循其成为普遍法则将使事情进展最好的原则，因为它们是其成为普遍法则是每个人都能够合理地意愿的仅有原则。

我认为，我们应当接受(S)和(T)，而有接受(R)和(U)的强理由。

或许值得解释一下，我为什么花那么多篇幅来辩护(U)。在怀疑存在道德真理的理由之中，最强的理由之一来自某些类别的道德分歧。大多数道德分歧对于存在道德真理的信念并不构成强烈的反对，因为这些分歧取决于不同的人持有相冲突的、经验的或宗教的信念，或者是他们有着相冲突的兴趣，或者是他们使用不同的概念，或者是这些分歧是关于临界个案(borderline case)，或者是它们依赖于所有问题都必须有答案或精确答案的错误假定。但是，有些分歧不属于上述类别。最深刻的分歧发生在这样的情况下：我们不是在考察特定行为的不当性，而是在考察道德与道德推理的性质，以及关于此类问题的不同观点蕴含着什么。如果我们和他人持有相冲突的观点，而我们毫无理由地相信，我们更可能是正确的，那就应该至少使我们怀疑自己的观点。它或许还给予我们以理由去怀疑是否有任何人可能正确。

419

一直以来人们广泛相信，康德主义者、契约论者，以及后果论者之间存在着这样的深刻分歧。我已经论证，情况并非如此。这些人是在不同的侧面攀登同一座山峰。

一直以来人们还广泛相信，不存在任何重要之事，因为理由是由我们的欲望给予的，而我们拥有这些欲望是毫无理由的。如我已经论证且将在第六部分进一步论证的，我们应当拒绝这种悲观的观点。

如今最重要之事在于，我们那些富人放弃某些奢侈的享受，不再让大气过暖，并在其他方面爱护这个星球，以便它继续供养有智慧的生命。如果我们是宇宙中唯一的理性动物，那么甚至更为重要的是，我们的后代在如其可能的几十亿年之间能否延存。也许我们有些后代会过的生活和创造的世界，虽然无法为过去的苦难辩解，但会让包括受苦者在内的我们所有人都有理由为宇宙的存在而高兴。

附录A 状态给予的理由

420 根据我们称之为的

状态给予的理论 (State-Given Theory)：只要某些事实使得我们如果持有某种信念或欲望就会更好，这样的事实就给予我们某种持有该信念或欲望的理由。

要确定我们是否有此类状态给予的理由 (state-given reasons)，可以先问问，我们可能怎样去回应这样的理由。

设想

情形一：某个反复无常的暴君发出可信的威胁说，我将要受十分钟的折磨，除非从现在起的一小时之内，我既相信 $2+2=1$ ，又想受折磨。某个测谎仪将揭示出，我是否确实有这样的信念和欲望。

按照状态给予的理论，这个人的威胁就给了我强烈的状态给予的理由去持有这种信念和欲望，因为这是我免受折磨的唯一途径。但我无法通过选择持有这种信念和欲望来回应这样的理由。

这里的一个问题在于，我持有对象给予的理由，它们决定性地反对相信 $2+2=1$ ，且反对想受折磨。姑且假设，由于没能持有这种信念和欲望，这个

暴君就折磨了我。有人可能会说：“笨蛋！你何不相信 $2+2=1$ ？”但这种说法很荒谬。我情不自禁地相信 $2+2$ 不等于 1 。我不想受折磨很傻，这样的说法也是荒谬的。我如果得知这是我取得某种大利的唯一途径，那或许会想受折磨。例如，如果我得了某种致命的疾病，而巨大的痛苦会启动我体内的治愈进程，那我或许会那么想。但该例不属于同种类型。这个暴君将兑现他的威胁，除非我想受折磨；但这不是作为某一目的的手段，而是作为一个目的或者说为了受折磨本身的缘故。我既然是理性的，那就不可能为受折磨本身的缘故而想受折磨。鉴于受折磨的可怕性，我就有决定性的、对象给予的理由，不去拥有受折磨的欲望，并且禁不住以这种非自愿的方式来回应这种理由。

接着假设，这个暴君给我一个容易一些的任务。

情形二：我将受折磨，除非从现在起的一小时之内，我相信某个封闭的盒子是空的。

按照状态给予的理论，这个威胁给予我一个状态给予的理由去持有该信念。而且这个理由不会遭到反对，因为我没有任何不相信这个盒子是空的、对象给予的认知理由。但如前所示，我无法通过选择持有这样的信念，来回应所谓状态给予的理由。我既然是理性的，就无法仅仅由于我有该信念会更好而选择相信该盒子是空的。

存在其他可能性。如果我们拥有某信念会对我们更好，此时主要有三种方法可以使得我们自己能拥有该信念。一种方法是使得该信念为真。例如，在情形二中，我或许能够打开密封盒，并取出其中装着的任何东西。那将使我相信这个盒子是空的，由此免于暴君的威胁。

在某些其他的情况下，我们可能致使自己拥有某一有益的信念（**beneficial belief**），其方法是通过找到证据或论证以给予我们持有该信念的、足够强的认知理由。这个方法有风险，因为我们可能找到的证据或论证会给予我们强烈的理由不持有该信念。但我们可以通过设法避免认识到这样的理由来降低风险。比如说，我们如果在努力相信上帝存在，那就可以去读信奉者的著作，并避免接触无神论者的著作。值得补充一点，在这么行动时，我们不仅在实践上而且在认识上都可能是充分理性的。对于我们意识到的任何认识的或显见的理由，我们可以总是合理地予以回应。我们必须小心避免意识到那些使我们不相信正在努力相信之事的认知理由，其原因或在于此。

在第三种情况下，我们如果持有某种自知为假（false）的信念就会更好，因为我们意识到给予我们不持有该信念的决定性的认知理由。我们如果是理性的，那就无法在意识到上述理由的同时做到持有该信念。但是，通过使用像自我催眠那样的技术手段，我们或许能够使自己持有之。我们不可能选择给予自己这样的信念：其内容使之显然为假。在暴君发出初次^①威胁时，我不可能使自己相信 $2+2=1$ 。无人能够在理解这个数学公式的同时相信其为真。但设想

情形三：该暴君威胁我将受折磨，除非从现在起的一小时之内，我相信他是世上最伟大的天才。

我或许能够通过自我催眠而持有这一错误信念。我将不得不让自己忘记，相信这个人不是天才的认知理由。我可能还必须让自己忘记，我是如何以及为什么使得自己持有这一新的错误信念，因为记得这些事实可能会削弱该信念。我既然是理性的，就不可能相信我知道毫无认知理由去相信的东西。由于类似的理由，我们可能还必须给予自己某些有关暴君辉煌成就的、貌似真实的假记忆（false apparent memories）。但我如果是个老练的自我催眠师，或许就能够做到这些事。由此我就会合理地逐渐相信，这个人是最伟大的天才，因为这些貌似真实的假记忆会给予我持有该信念的、决定性的显见理由。

我们大多数人并没有这样的自我催眠能力。但我们可以设想逐步拥有它们。由此我们就能够使得自己随意拥有许多的错误信念，它们可以像我们进行其他各种心理活动一样来得即时迅捷。

现在回到我们可以持有状态给予的理由这一观点。状态给予的理论家们主张

（1）只要某些事实使得我们如果持有某种信念就会更好，这样的事实就给予我们某种持有该信念的理由。

在我刚才描述的各种情况下，我们毫无诉诸如此理由的必要。如下主张就足够了：我们有理由希望持有这样的有益信念，并且尽可能致使自己持有之。这就像希望某事发生并尽可能使之发生的任何其他理由。添加如下说法是毫无

^① 指情形一。以下的第二次、第三次依此类推。

意义的：我们不仅有理由致使自己持有这样的信念，而且有理由持有之。

可以设想出我们心理上的另一种变化。如下情况可能变成真的：在相信我们如果持有某种认识上不理性的信念就会更好时，我们有时并不需要用像自我催眠这样的、有意的心理活动来使自己持有该信念。以某种非自愿的（non-voluntary）方式通过若干组支持性的、貌似真实的假记忆，我们可能不自觉地（find ourselves）开始拥有这样的有益信念。

或许看上去有道理的是：在这样的情况下可以大肆宣扬，我们拥有状态给予的理由去持有此类信念。如我所言，在意识到给予我们拥有某信念的、决定性的认知理由时，我们会回应这样的大多数理由，其回应的方式不是通过有意致使自己持有该信念，而是通过一种非自愿的方式开始拥有并持存该信念。我们可能做出类似的主张：在不自觉地开始拥有此类不合理但有益的信念时，我们就是在对拥有该信念的实践理由做出回应。

我建议，我们应当拒绝上述主张。还有两种更好的方式来描述这种情况。

按照一种描述，就开始拥有这些有益信念而论，我们尽管是以某种非自愿的方式，但仍然是在回应“致使自己持有该信念”的理由。我们经常不自觉地做某种我们也能够自愿（voluntarily）做的事。例如，我们可能不自觉地突然想努力抓住某个我们刚掉落的东西，或移动身体以恢复平衡，或者是在身体下落时举起手臂以护住脑袋。如果看到某个手榴弹就要爆炸，我们可能不自觉地扑到手榴弹上，以救身边人的生命。这些都是无意之中对于按照某些方式行动的理由而做出的回应。姑且假设，在暴君发出初次威胁时，我不自觉地开始相信他是个天才。在此，我可能就是在以这种非自愿的方式，对于致使自己拥有该有益信念的实践理由做出回应。这可能就是在某些不自觉地自我欺骗的实例之中发生的情况。

424

我们可能转而主张：在不自觉地开始拥有这样的有益信念时，我们并不是在回应任何理由。真相或许仅仅在于，在相信我们如果拥有某一其他信念就会更好时，这种相信就致使我们去拥有这个其他的信念。这有些类似于如下方式：在相信我们处于危险时，这种信念致使肾上腺素在我们的血液中释放，从而有助于我们更有效地应对该危险。在此肾上腺素的释放虽然是有益的，但并不涉及对某种理由的回应。某些一厢情愿的情况或许也不涉及对理由的回应。

现在回到这一主张：在这样的情况下，我们是在对拥有此类有益信念的理由做出回应。我已提出，我们应当拒绝该主张。如果我們是在致使自己拥有此类信念，这个过程就可能是合理的，且包含对理由的回应。我们是在对行动的

理由做出回应，该理由产生于“我们如果拥有此类信念就会更好”这一事实。但是，如果我们仅仅是被动地（*passively*）开始拥有此类信念，这个过程就不会是合理的或涉及任何对理由的回应。假设我无法自我催眠去相信该暴君是天才，其结果是受到了暴君的折磨。有人可能会讲：“笨蛋！你为什么不对相信这个人是天才的理由做出回应？”当意识到某一事实给予我们决定性的认知理由要拥有某个信念时，我们如果没有回应这种理由而拥有该信念，那就不是完全理性的。但如果无法致使自己拥有某种有益但不合理的信念，那么我们即使没能拥有该信念，也不应经受哪怕丝毫的指责。而即使不能对我们意识到某些所谓理由做出回应，我们也绝非不理性的；若如此，则可算作是反对这一观点，即我们拥有诸如此类的某种理由。

425 我们还有其他理由去拒绝状态给予的理论。我们可以说，两种理由是

互竞的（*compete*），如果我们不能对这两种理由同时成功地做出回应。

以及它们是

冲突的，如果对于同一个问题，它们支持不同的答案。

例如，如果我们拥有守诺的道德理由，同时又有违背诺言的自利理由，这两种理由就是互竞的，因为我们不可能既守诺又违背诺言。它们也是冲突的，因为对于我们最有理由做什么这一问题，它们支持不同的答案。

接着假设我们意识到，有些事实给予我们决定性的认知理由不拥有某种有益的信念。根据状态给予的理论，拥有该信念带来的益处，也会给予我们状态给予的理由要拥有它。这两组理由是互竞的，因为我们不可能既拥有又不拥有该信念。按照这种观点的一个版本，它们也是冲突的。在问我们最有理由相信何者时，这两种理由对于同一个问题支持不同的答案。我们将不得不就此做出决断：状态给予的理由让我们拥有该信念，认知理由让我们不拥有该信念；这两种理由孰强孰弱。

我认为，我们并不会拥有如此相冲突的理由。在暴君做出第三次威胁时，我将意识到这样的事实：它给予我决定性的认知理由，让我不会误以为这个人是世上最伟大的天才。如果我拥有状态给予的理由去拥有这一信念，那么这种

理由会产生于受折磨会很糟这一事实。我可能这样问：与受折磨相比，我拥有如此的错误信念是否会更坏。但在此我问的是：在这两种结果之中，我更有理由想要并努力去阻止哪一种。这是一个有关两种实践的理由的强度问题，它们类似于任何想要和努力去阻止某种坏结果的其他理由。我不能合理地问道，我拥有这种错误信念的、状态给予的理由是强于还是弱于不拥有该信念的认识的理由。比较证明该信念的虚假性之证据强度与我受折磨的坏处，这是毫无意义的。

既然意识到这样的比较毫无意义，状态给予的理论家可能就转而主张：这两种理由并不冲突，因为它们所支持的是对于不同问题的回答。在问我们是否应当拥有某信念时，我们在问的可能要么是

Q1：这是一种我们应当在认识上拥有的信念吗？

要么是

Q2：这是一种我们应当在实践上拥有的信念吗？

按照这种观点，在回答Q1的过程中，我们应该仅考虑认识的理由；而在回答Q2的过程中，我们应该仅考虑实践的、状态给予的理由。既然它们是不同的问题，那我们就不能这样问：基于全面的考虑，我们应当或最有理由相信什么。

上述主张部分正确。这里确实存在两种问题。但该主张无助于说明，我们可以有状态给予的实践理由去拥有信念。Q2有待解释，因为我们是否应当在实践上拥有某信念，这一提问的意思是不清晰的。我提议，这个问题可以更清晰地陈述为

Q3：对我来说相信什么会最好？换句话说，我最有理由想要相信什么，并且尽可能地致使自己去相信？

而这个问题并非关于我有理由相信什么。像其他的实践问题一样，该问题是关于我有理由想要并且去做什么。

既然Q1与Q3是不同的问题，那我们就绝无必要去比较实践理由与认知理由这两者的强度。我们对理由做出回应。而我们永远也不可能既拥有认知理由不以某种方式做出回应，同时又拥有实践理由要以这种方式做出回应。当暴君

做出其第三次威胁时，我可能以这样一种方式来行动——使自己相信这个人是最伟大的天才——由此对我的实践理由做出回应。我毫无不这么做的认知理由，因为认识的理由并不是行动所依据的理由。我的确拥有不相信这个人是如此天才的、决定性的认知理由；而在记得给予我这些理由的事实时，我可能通过丧失该信念而以一种非自愿的方式对这些理由做出回应。但我毫无不以这种非自愿的方式去回应的实践理由。我的实践理由是：按照会使我保持“相信暴君是天才”这一信念的方式行动，直至通过暴君的测谎检验，从而他不会折磨我。对于这些实践理由与认知理由，我不可能都成功地做出回应；在此意义上，这两组理由是互竞的。但它们并非冲突的。

我所做出的如上区分，易为人忽视或误解。如我所言，理论推理是一种自愿的活动，它们通常是我们出于实践的理由而进行的。例如，在做数学运算时，我们可能拥有某种实践的理由，去检查某一部分的证据，或以另一种方式重新运算。这些是以可能有助于我们达到真理的方式行动的理由。我们出于实践的理由而这么做，同时也要回应诸多认识的理由。例如，在检查某一证据时，我们一旦看出由某某得出某某或何者正确，那就回应了认识的理由。开始拥有如此这般的特定信念，这并非一种自愿的心理活动。可以说，理论推理既包含实践的合理性，也包含纯粹认识的合理性。

428 实践理由与某些认知理由之间还存在其他的密切关联。我们的许多实践推理在于有关实践问题的理论推理。在问我们最有理由做什么时，我们或许是在努力探究对这个问题的某种正确答案。而有些事实可能既给予我们决定性的实践理由去按某种方式行动，也给予我们决定性的认知理由以相信我们拥有这种实践理由。回到这个情形：你的旅馆失火了，你只有跳进某条河才能逃命。这一事实给予你跳河的决定性理由，也给予你相信你应当跳河的决定性理由。但尽管我们的实践理由与认知理由通常是紧密相关的，且此类理由可能是互竞的，但它们永远不可能冲突。

状态给予的理论家们还主张：

(2) 只要某些事实会使我们如果拥有某种欲望就会更好，这些事实就给予我们拥有该欲望的理由。

与我们可以拥有状态给予的理由去持有某信念这一主张相比，主张(2)看上去要有道理一些。我们可以反驳说，既然信念旨在真实(truth)，我们拥

有信念的理由就必须全是认识上的或与真实相关的。这样的反驳绝不适用于欲望。因此，如下说法或许看上去有道理：正如我们有对象给予的理由去拥有某种欲望，当且因为我们想要的事物（**what we want**）在相关方面是好的；我们也有状态给予的理由去拥有某种欲望，当且因为我们对某事物的想要（**our wanting something**）是好的。

我要提出，我们并没有这样的理由。假设

情形四：我的暴君宣告：我将要受十分钟的折磨，除非在接下来的一小时之内，我想要他杀死我。如果我拥有该欲望并请求他杀死我，那他就会拒绝并且放过我。就我所知，这个人总是言出必行。

进而假设我的余生会活得相当有价值。由此我觉得很难想要他杀死我。但我或许能够自我催眠，在随后的几小时之内拥有该欲望。这将是我最有理由去做且合理地应当做的事。这种心理活动将是避免强烈痛苦的安全方法。

状态给予的理论家们可能宣称，他们的观点解释了我为什么应当这么做。他们可能论证道：

(A) 我有想要暴君杀死我的决定性理由，因为这将使我免受折磨。

(B) 当我们有决定性的理由去拥有某种欲望时，这一事实给予我们促使自己拥有该欲望的决定性理由，只要我有某种这么做的安全方法。

(C) 我有这样的方法来使自己想要他杀死我。

429

因此

我应当使自己拥有该欲望。

然而，前提(A)为假。我有想要暴君不杀我的、对象给予的理由，而这也是不想要他杀我的理由。这些理由显然强于我所谓想要他杀我的、状态给予的理由。失去值得过的人生比受十分钟的折磨要坏得多。因此，我并没有希望他杀我的决定性理由。

状态给予的理论家们可能答道，我并没有任何不想要这个人杀我的理由。

如果我拥有该欲望，这个人不会杀我，而是放过我。既然我有理由去拥有该欲望，且毫无不拥有它的理由，那我就应当合理地使自己拥有该欲望。按照这种观点，拥有欲望的所有理由都是状态给予的，或者是由拥有这些欲望的益处而产生的。

要评估这一观点，我们可以假设，我拥有该欲望的这种努力归于失败，暴君因此折磨我。有人可能说：“笨蛋！你为何不想要他杀你？”但这是无稽之谈。如前所示，我如果是理性的，就不可能仅仅因为我知道如果有该欲望就对我来说更好而想要这个人杀我。用一个更简单的例证可以更清楚地说明这一点。如果我得知自己患有致命疾病，我如果想要死去对我来说可能更好。但这并不说明我毫无不想死的理由。他人这么说就是荒唐的：“笨蛋！你为何不想死？”应该承认：即使在暴君发出威胁之后，我仍然有决定性的、对象给予的理由想要他不杀我。

状态给予的理论家们接下来可能提出：这些理由既然属于不同的类型，那就不相冲突。按照这种观点，我们可以提两个问题：

Q4：我有最强的、对象给予的理由想要什么？

Q5：我有最强的、状态给予的理由想要什么？

430

但该提议不成功。我们也可以问：

Q6：基于全面的考虑，我最有理由想要什么？

如果对于拥有同一欲望来说我们有支持和反对的理由，那么这些理由就的确相冲突，因为对于这个更宽泛的问题，它们支持不同的答案。在此不相干的是，这些理由属于不同的类型。可以类似地主张，道德的理由与自利的理由属于不同的类型——但是，在问基于全面的考虑我们最有理由做什么时，这些理由由于对该问题支持不同的答案而可能发生冲突。

我们现在讨论的这类情况之中，有两个值得问的问题。但这两个问题不是关于支持或反对拥有同一欲望的两种理由。Q6可重述为

Q7：我最有理由拥有哪些欲望？

我们还可以问：

Q8：我最有理由想要拥有哪些欲望，并且尽可能地致使自己拥有？

在情形四中，我可以问：

我如果想要暴君杀我，那是在想要某种“我有决定性的理由想要”的事吗？

我如果致使自己拥有该欲望，那是在做某种“我有决定性的理由去做”的事吗？

我的答案应该分别是否定和肯定的。我如果想要这个人杀我，那这一欲望本身是不合理的，因为我拥有“不想要他杀我”的决定性理由。但致使自己短暂地拥有这一不合理的欲望，对我来说是理性的，因为这种行动会让我免受折磨。

还有另一种情况可以让我们有理由拒绝承认：我们有状态给予的理由去拥有欲望。设想

自我挫败的欲望（self-defeating Desire）情形：我有入睡的强烈欲望，因为我需要睡眠以增进我明天面谈的表现。但我有某种失眠症。只要我强烈地想要入睡，这一欲望就让我担心自己睡不着，由此我会保持清醒。因此，我只有放弃入睡的欲望才能入睡。

431

我需要睡眠，这给予我对象给予的理由想要入睡。根据状态给予的理论，该需要也给予我“不拥有这一欲望”的状态给予的理由，因为这是我入睡的唯一途径。这些理由是相冲突的，因为它们分别是对于拥有同一欲望来说支持与反对的理由。按照这种观点，要决定我是否应当拥有该欲望，我就应该比较这两种理由的强度。我应该问：基于全面的考虑，我最有理由想要什么。

我可以很容易地比较这两种理由的强度。我想要入睡的、对象给予的理由产生于这一事实：我需要睡眠以增进我明天面谈的表现。我所谓不拥有该欲望的、状态给予的理由产生于同样的事实，加之以拥有该欲望会让我保持清醒这一事实。既然这两种理由都从我对睡眠的需要这里获得其规范力量，它们的强度就正好相等。它们既然又是相冲突的，那就会相互抵消。因而状态给予的理

论就蕴含着：总的说来，我没有任何想要入睡的理由。如果这一点正确，我就毫无理由拥有入睡的目标，并且毫无理由去致使自己放弃这一欲望以便实现该目标。这些断言显然为假。

432 我的提议是，我们应当拒绝这种状态给予的理论。我没有任何状态给予的理由不去拥有入睡的欲望。我所拥有的是对象给予的理由，由此想要没有该欲望，且尽可能地致使自己放弃该欲望。不同于我所谓“不拥有该欲望”的状态给予的理由，此类理由并不冲突于“我拥有该欲望”的对象给予的理由。按照这种观点，我们会达至正确的结论。我的睡眠需要，给予我不受反对的强理由拥有“想要入睡”的欲望；该需要也给予我不受反对的强理由不拥有该欲望，因为这是满足该欲望的唯一方式。

只要我们如果拥有某些信念或欲望就会更好，我们就有理由想要拥有这些信念或欲望，且尽可能让自己拥有。但我要提出，我们并没有状态给予的理由去拥有信念或欲望。

我们或许有状态给予的理由而处于某类其他的状态。例如，我满可以这样宣称：我有某种理由明年四月呆在巴黎。但如我所论证的，这样的理由微不足道。如下的说法即已足够：我有理由想要明年四月在巴黎并尽可能去那儿。

附录B 合理的不合理性与高契尔的理论

在一篇早期论文中，高契尔坚决主张：我们要合理地行动，就必须道德地行动。我力图驳斥这种论证。由于未能说服高契尔，我要再试一次。 433

高契尔认为，我们要是理性的，就必须使自己的预期效用最大化。尽管他在“效用”与“收益”（benefit）之间做了区分，但该区分并没有改变其主要论点。我们可以认为他是在诉诸合理利己主义。

许多作者已主张，在自利的意义上（in self-interested terms），道德地行动总是合理的。根据其中大多数作者的看法，道德与自利是一致的。但这并非高契尔的思路（line）。高契尔承认，道德地行动可能对我们更不利，且这一点已为人所知。他主张，即便在这种情况下，道德地行动仍是合理的。

如果我们诉诸合理利己主义，那么为这种主张辩护看来就是不可能的。如果知道我们的行动对我们更不利，那么在自利的意义上它们如何可能是合理的？但是，高契尔修正了合理利己主义。按照该理论的标准版，某一行动是合理的，如果它使我们的预期收益最大化，或对于我们来说是可预期地最佳的（expectably-best）。按照高契尔的版本，合理的不是用我们的行动去自利，

而是用我们的习性^① (dispositions) 去自利。如果拥有某一习性对我们来说是预期最佳的, 那么该习性就是合理的。就做出如此主张而论, 高契尔的观点类似于某个版本的间接后果论。

除修正合理利己主义之外, 高契尔还限制了道德的范围。他主张, 我们要道德地行动, 就必须尊重我们的协议。在其所关注的情况之中, 我们每一个人都承诺, 我们将付出一定的成本, 以给予他人更大的收益。我们如果全都遵守这样的承诺就会全赢。每一个人所付出的成本, 将为每一个人从他人那里获得的更大收益所超过。

尽管这样的协议是互利的, 但对于我们每一个人来说如果他/她不守诺通常会更好。我们无论是暗中违背还是不惜损害名声, 其代价都会小于我们由此之所得。我们可能认为, 在自利的意义上, 违背这样的承诺就是合理的。但高契尔坚决主张, 我们如果这么做, 那就是傻瓜。

高契尔的论证始于一个预测。我们如果是直截了当地自利的——或简言之, 精明的 (prudent) ——那就会图谋违背这样的承诺。其他人得知这一点, 就会把我们排除于这些互利的协议之外。这对我们更不利。我们如果是可信赖的 (trustworthy), 那对我们会有利, 因为由此我们就为这些协议所接纳。

如我所指出的, 我们如果只是显得可信赖但实际上是精明的, 那对我们来说甚至更有利。我们仍会为这些协议所接纳, 但只要可以预期违背承诺对我们有利, 我们就会这么做。高契尔的答复是: 我们都是可识破的 (translucent), 乃至没有如此的欺骗能力。在订立这样的协议时, 我们有时不能隐瞒自己的真实意图。因而他宣称, 总的说来, 如果我们是真正可信赖的, 那就对我们更有利。

接着高契尔诉诸其合理利己主义的变体——我将称之为高契尔观点 (Gauthier's view)。按照该观点, 成为可信赖者既然符合我们的利益, 那么

① Disposition目前有多种中译, 不统一。在哲学伦理学文献中, 大概有“气质”、“意向”、“取向”等译法; 经济学与博弈论中另有“处置”、“外征”等译法。disposition为本篇附录的关键词, 有必要统一译法以免混乱。但单取原有的某一种译法不太吻合本篇的文义: 其一, 它是后天习得的 (acquired), 有“后天形成的行为气质”的意思。其二, 有习惯做法的意味。其三, 不止于意向与气质, 比意向更趋向于实践, 其重心在于对行为表现的支配。在此, 译为“处置”、“外征”则失去第一、二点所要表达的意思。“取向”则不足以表达第二点与第三点。综合各方面的考虑, 本人认为译为“习性”可能更接近该文要表达的意思。但这种统一译法仅限于本篇, 本书其他各处的译法将随语境而调整。

按此习性行动对我们来说就是合理的。守诺是合理的，即便是在我们知道在做之事将对我们更不利时也是如此。

我们应该接受这种论点吗？我认为不。在运用于可信赖（*trustworthiness*）时，这种论点或许看上去有道理。但我们应该拒绝高契尔观点。拥有某种习性可能符合我们的利益，且使自己拥有该习性是合理的，但依之行动则是不合理的。

二

高契尔观点的一个难题在于，在不同的时候，符合我们利益的习性也不同。由此很难用可能实现其目标的方式来陈述高契尔观点。 435

在对其观点的最初陈述中，高契尔认为

（A）如果我们养成了某种习性——因为我们合情理地相信，这么做会让我们的生活更好——那么依此习性行动就是合理的。

我对（A）的质疑如下。正如成为可信赖者可能符合我们的利益一样，倾向于执行我们发出的威胁而无视他人的威胁，也可能符合我们的利益。如前所示，最有利的行动会是：在实际上保持精明的同时显得拥有这样的习性。但为了检验高契尔观点，我们应该接受他的主张：我们都是可识破的乃至不能欺骗他人。那么对于我们来说可能还是确实拥有这样的习性为好。但是，依此习性行动对我们来说可能不是合理的。

我在此将如下例子称作你的致命威胁情形（*Your Fatal Threat*）。假设你我共处荒岛，我们都是坦率的（*transparent*）。你成为一名威胁执行者（*threat-fulfiller*）。通过经常威胁要引爆某个炸弹，你旨在迫使我成为你的奴隶。我维护自由的唯一途径是成为一名无视威胁者（*threat-ignorer*），倾向于决不妥协于你的威胁。我既然是可识破的，就可以合情理地预期你会意识到我的习性，由此对我最有利。我设法养成了这种习性。但我的运气不佳。在某次一时疏忽之中，你发出威胁：除非我给你一个椰子，否则你就将我俩都炸死。根据（A），对我来说无视你的威胁就是合理的。即便在如下情况下也是合理的：我知道如果我这么做，你就会引爆炸弹，将我俩都炸死。

高契尔一度接受这一结论。但他后来修正了他的观点，从（A）转向

(B) 如果我们有理由相信：养成某种习性让我们的生活更好，那么依此习性行动就是合理的。

436

根据(B)，按照某一习性行动要是合理的，这样就是不够的：我们以前的确有理由相信，养成该习性会让我们的生活更好。我们还必须相信，这个以前的(past)信念是正确的。我们没有必要“在面对这样的情况——已知某一习性不能让生活更好——之下，仍然坚持该习性”。

高契尔意图用(B)来应对我的例子。在你发出致命威胁时，我就丧失了持有如下信念的理由：成为一名无视威胁者会让我的生活更好。按照高契尔经修正的观点，我就没有必要“坚守”我的习性。

我们可以修改这个例子。假设我知道，我要不是成为一名无视威胁者，早在此前的某个时候就已死去。高契尔观点就又蕴含着：我应该无视你的威胁。既然我的习性曾救过我的命，那养成这种习性就让我的生活更好。的确，该习性现在会让我丧命。但这并非关键所在。根据(B)，我应该拒绝给你椰子而被炸死。

如此例所示，即便某一习性已成为灾难性的，(B)仍然可能蕴含着依此行动是合理的。如果该习性过去带来的收益大于其未来成本，这就是合理的。高契尔主张，我们应该“坚守”这样的习性。我们应该忠实于我们的“承诺”。

在运用于承诺时，这种观点有某些吸引力。我们如果受益于可信赖，就可能认为依此习性行动是合理的，即使它成为负担时也是如此。在此讨论承诺是有意义的。但就威胁行为而论，它就几乎没有意义。为什么我应该仅仅由于该习性救过我的命，就不惜以生命为代价而继续做一名无视威胁者？

如果我的另一个选项是成为你的奴隶，那么我的死亡很难说是一种代价。但我们可以给这个情形再增加一个细节。假设一个救援队刚刚登陆该岛。我知道，如果给你椰子，我不久就会自由。

437

为了应对情形的这一版本，高契尔必须再次改变其观点。对我来说或许原本合理的是成为一名无视威胁者。但如高契尔必定会同意的，现在对我来说合理的是设法放弃这一习性。如果我能够使自己放弃该习性，那么让自己保持该习性就是不合理的。既然如此，高契尔就不可能主张，依此习性行动必定仍是合理的。既然我不久就能获自由，故意招致杀身之祸对我来说就是不合理的。

高契尔应该怎样修正其观点呢？他或许重申主张(B)，以涵盖临时的习性。但有一种更简洁的表述方式。高契尔可以求助于

(C) 如果我们有理由相信：拥有某种习性将会让我们的生活更好，那么依此习性行动对我们来说就是合理的。

高契尔的观点如果诉诸(C)，那就不会受到我的例证的挑战。在我看出我的习性已变成灾难性的时候，(C)并不蕴含着：对我来说依之行动必定仍是合理的。

我再举一例，在此我称之为先令的情形(Schelling's Case)。一名抢劫犯威胁说，除非我打开保险箱，把所有的钱都交给他，否则他就要动手杀我的孩子。对我来说无视他的威胁是不合理的。但是，即使我屈服于他的威胁，仍然存在这样的危险：为了降低被抓的可能性，他把我们全都杀死。我主张，在这种情况下，对我来说合理的是服用一种使我变得极其不理性的药物。这个抢劫犯由此就会判断出：威胁我将毫无意义；而且既然他无法实施罪案，而我也无能力去报警，那他也不太可能杀我或孩子。

在考察这一情形时，高契尔似乎接受(C)。他同意，让我自己短期内精神失常对我来说是合理的；并且他主张，依此习性行动对我来说是合理的。

然而，如果求助于(C)，高契尔就要付出代价。在他为契约道德的辩护之中，高契尔所比较的仅仅是持久的习性。他认为说明这一点就够了：如果我们是可信赖的，那么总体上说这会让我们的生活更好。但如果诉诸(C)，他就需要做出更多的说明。根据(C)，按照某一习性行动要是合理的，仅说明养成该习性此前符合我们的利益就是不够的。我们必须有理由相信，在行动的那一刻，拥有该习性符合我们的利益。因而高契尔必须说明，如果我们是可信赖的，那么在我们信守协议时该习性符合我们的利益。 438

我认为，他并没有说明这一点。他至多只说明：在商定协议时可信赖符合我们的利益。在某些情况下，在要求信守一个协议的时刻，我们正在商定某个新协议。此时高契尔的论点可能适用。但在其他情况下没有像这样的吻合。有些承诺是我们可以秘密而迅速地违背以符合自身利益的。在可能出现这种情况时，我们如果是可信赖的就对我们更不利。如果我们放弃这一习性而变成自利的——即便这是暂时性的，仅足以违背承诺——那对我们更有利。

要辩护其观点即道德地行动总是合理的，高契尔必须主张，信守上述承诺是合理的。然而，如果诉诸(C)，他就要放弃支持该主张的论证。(C)蕴含着：违背这样的承诺是合理的，因为我们由此就是在按照这样的习性行动：我们能够合情理地相信该习性在当时对我们最有利。

高契尔可能试图给出另一种答复。他可能主张，我们如果是可信赖的，那就没有能力去放弃或克服这一习性。在与此处问题相关的意义上，该主张可能不正确。但姑且假设它是正确的。假设我是可信赖的，因而发觉违背承诺是不可能的。高契尔可以诉诸“应当”蕴含着“能够”这一主张。他可能会说，我既然无法违背承诺，那么如下情况就不可能为真：违背承诺对我来说是合理的。而且他还可以说，鉴于我的习性之强度，依此习性而行动对我来说就是合理的。

这个答复充分吗？回到我倾向于无视你的致命威胁这一情形。我如果克服我的习性，且由此设法保命直至可获救援，那高契尔必定同意，我的行动是合理的。但假设我的习性过于强劲。我发现无法使自己给你椰子。既然我无法克服我的习性，那对我来说这么做是合理的这一点就不可能为真——高契尔能够这样宣称吗？既然对我来说，做出有所不同的行动在因果的意义上是不可能的，那么我招致杀身之祸就是合理的——他能做出如此主张吗？

439 我认为答案是否定的。依据我前面给出的理由以及高契尔在别处的主张，对我们来说怎么做是合理的，就此而论并不取决于因果意义上的可能之事。在相关的意义上，我们本可以采取别样的行动，只要除我们的欲望或习性之外别无阻碍我们这么做的其他因素。如果对我来说原本采取别样行动是合理的，那么如下情况就是不相关的：鉴于我的欲望和习性，别样的行动在因果意义上原本是是不可能的。我也不能诉诸我的习性之强度来为该行动辩护。这可能让我免于某些类别的批评，但无法表明我的行动是合理的。

高契尔放弃了主张（A），就此而论，他也承认我的上述观点。假设尽管养成某一习性在过去对我来说是合理的，但我已得知这么做是一个可怕的失误。高契尔就不再主张，按照这样的习性行动必定仍是合理的。他同意，我们无法从这样的事实——我合理地养成某一习性以及我现在无法克服该习性——推出：按照该习性行动对我来说是合理的。

三

我已说明高契尔观点的一个难题。既然拥有临时的习性可能符合我们的利益，那就很难用可能实现其目标的方式来陈述高契尔观点。现在让我们忽略这一难题，而转向核心的问题。我们应该接受高契尔观点吗？如果拥有某一习性符合我们的利益或使自己拥有该习性是合理的，那么依之行动就是合理的——我们应该相信这一观点吗？

在我们所关注的情形之中，尽管拥有某一习性符合我们的利益，但依之行动则违背我们的利益。唯有在此高契尔观点才重要。

重新考察先令情形。由于我暂时的精神失常，抢劫犯知道：他即使动手伤害我的孩子，也不会由此诱迫我打开保险箱。这给予他放弃并离开的理由，由此对我大为有利。但是，当我处在受药物控制的状态，且在抢劫犯离开之前，我是以破坏性的和自我挫败的方式行动。我由于爱而打孩子，由于想保护我的手稿而烧掉它。

高契尔反对说，我的疯狂行动实际上对我有利。它们是在说服这个人，其威胁对我不起作用。既然这样的行动对我更有利，那从任何观点看它们都是合理的。因此，这并非如我所主张的是一种合理的不合理性之情形。 440

为答复这一反驳，可以对该情形增加一个特征。我们可以假设，要让这个人相信我处在疯狂状态，我没有必要以疯狂的方式行动。他看到我服药，并且知道它导致暂时的疯狂。既然抢劫犯已知道我处在这种状态，那么我的破坏性行动就毫无好效果。

我的行动尽管只有坏效果，但产生于一种有利的习性。按照高契尔的观点，这就足以使这些行动成为合理的。

休谟有个著名的主张：我们宁愿自己彻底毁灭，也不愿引起某个陌生人的丝毫不安，这并不违反理性。但高契尔的观点更为极端。休谟至少要求，我们的行为要是合理的，我们必须是在努力实现自己的目标。按照高契尔的观点，我们可以是在努力挫败自己的目标。在烧我的手稿或打我的孩子时，我是在做我相信是不合理之事，且因为我相信它是不合理的。我的行动极尽疯狂之能事。按照高契尔的观点，它们仍然可能是合理的。这显然是错误的。

四

高契尔有一个支持其观点的论证诉诸这样的主张：如果我们接受他的观点，那对我们更有利。我们可以先问问这是否正确。

高契尔认为，我们要是理性的，就应该让我们自己的预期效用最大化。他将这一观点的两种版本做比较。根据合理利己主义的标准版（可称之为E），我们应该在行为的层次追求最大化。某一行为如果让我们的收益或预期收益最大化，那就是合理的。根据高契尔的观点，我们应该仅在习性的层次追求最大化。某一行为如果产生于使收益最大化的习性，那就是合理的。在此可将该观

点称为G。

441 在我们所关注的情形之中，我们无法在这两个层次都始终使预期收益最大化。我们如果力图就我们所有的行为来追求最大化，那就无法拥有使收益最大化的习性。因此，如果我们只要预期违背承诺对我们更有利就这么做，那就无法成为可信赖的，可信赖会对我们不利。

我们无法在两个层次都追求最大化，此时如果拥有收益最大化的习性对我们来说会更有利。这种习性的良好效果将超过我们行为的不利效果。

高契尔主张，鉴于这一事实，我们如果接受G而非E就对我们更有利。在做出这一主张的过程中，高契尔假设，我们如果接受E，就会以行为而不是习性来追求最大化。

该假设可能不正确。既然如果我们拥有最大化的习性就对我们更有利，E就会建议我们尽可能养成这样的习性。E认同于G的观点：我们应该努力拥有这些习性。E所拒绝承认的，仅仅是按照该习性行动必定是合理的。

高契尔可能认为，我们如果接受E，就总是会做E主张为合理的事。或者他可能认为，就判断任何关于合理性的理论而言，我们应该问的是：如果我们始终成功地遵循该理论，那将是怎样的。他假设我们总是以行为来追求最大化，其原因或在于此。但我们如果改变我们的习性，那就无法总是做E主张为合理的事。养成这些习性，其本身就是一种最大化的行为。我们如果以所有其他的行为来追求最大化，那么就养成这样的习性而论，我们的行为就是不合理的。我们如果转而养成这样的习性，那就无法总是以其他行为来追求最大化。

既然无法总是做E主张为合理的事，我们就必须做尽可能最有利的事。而E蕴含着：与其以我们的其他行为来追求最大化，我们不如养成使收益最大化的习性。这是我们可以预期对我们最有利的行动方式。E与G之间的分歧不在于我们是否应该养成最大化的习性这一问题。与G一样，E主张我们应该养成这样的习性。分歧仅仅是关于，我们在按照该习性行动时所做之事是否合理。

442 现在高契尔可能会说，如果接受E，我们就不能养成这样的习性。我们会认为，在某些情况下，按照这样的习性行动是不合理的。而我们或许不能使自己有意图去做我们认为是不合理的事。也许为了养成这样的习性，我们必须接受高契尔的观点，并且相信按照这样的习性行动是合理的。

在讨论核威慑（nuclear deterrence）的时候，高契尔的确做出了这样的主张。他设想，如果我们受到攻击，那么产生报复的意图将符合我们的利益。产生该意图可能保护我们免受攻击。高契尔接着声称，如果相信这样的报复是不

合理的，我们就不能产生该意图。

如果相信依此习性行动是不合理的，我们就绝不可能养成这样的习性；这样的主张是没有道理的。先令的情形是一个例外，还存在许多其他的情况。但高契尔并不需要这么强的主张。他可能说，养成这样的习性通常是不可能的；或者说，如果我们相信依某种方式行动是不合理的，那我们要变得有意这么做就更困难。我们可能不得不使用某种间接的方法，比如服药或催眠，这两种方法都有害处。如果我们相信这么做是合理的，那事情就可能容易一些。我们或许由此能够径直决定就这么做。

这可能只是转换了问题。我们如何可能获得这种信念？假设如高契尔所主张的：我们不可能意图报复，除非我们相信报复是合理的。如果如高契尔所承认的，报复既无意义也是自杀性的，那我们如何可能说服自己——亦如高契尔所主张的——这样的报复是合理的？我们如何能够让自己相信高契尔的观点？就获得某一信念而言，如果其唯一根据是该信念符合我们的利益，那我们并不容易做到。在此，我们可能也需要某种代价高昂的间接方法，但姑且置之不论。接着假设获得某种有益的习性对我们来说是不可能的，除非我们可以设法相信依之行动是合理的。由此使自己获得这种信念就可能符合我们的利益。那么，接受合理利己主义的标准版本就对我们更差，而接受高契尔的观点对我们更好，但这仍然没有说明：高契尔的观点是正确的，或者是最佳的。要达成这一结论，高契尔还需要另一个前提。

443

在其论证的最初版本中，高契尔的另一个前提是合理利己主义的标准版，这一点令人惊讶。他认为，我们应该从接受E入手。我们应该相信，某一行为如果对我们预期最好就是合理的。接着他主张，如果我们通过从E转到G而改变自己关于合理性的观念，那对我们更好。既然我们如果做出这一改变就对我们更好，E就蕴含着这么做是合理的。S^①要求我们相信，正确的理论不是E而是G。高契尔断定，正确的理论是G。

卡根提出了如下反驳。若E为真，则G必假，因为E与S不相容。若E为假，则G可能为真，但是，“E要求我们相信G”这一事实并不支持G。一种假的理论要求我们去相信某事，这是不相干的。无论以其中哪一种方式，高契尔的论证都不能支持其结论。

① 当为自利理论 (Self-interested Theory) 的缩写。

高契尔后来修正了他的论证。他不再主张，我们应该先接受E，然后转到他的观点。他直接坚持认为，我们应该接受他的观点。

在其论证的这一版本之中，高契尔的主要主张似乎仍然是：如果我们接受他的观点，那对我们更好。他的另一前提应该是什么？

尽管不再诉诸E，高契尔可能仍会说，如果接受某一信念符合我们的利益，那这么做就是合理的。由此他可能坚持他的这一主张：接受G对我们来说是合理的。

如前所示，这样的主张并不蕴含着G为真。接受一种假的理论也可能是合理的。但高契尔可能认为，说明接受其观点是合理的就足够了。他可能会说，即使在科学之中，我们也无法证明我们的理论是真的。我们至多揭示，相信它们是合理的。

然而，如此论证会混淆两种类型的合理性。在主张拥有某信念是合理的之时，我们通常意指，该信念在理论上或认识上是合理的，因为我们有充足的认知理由去拥有。这样的理由支持该信念，因为它们产生于这样的事实：该事实
444 或是使该信念成为必然的，或是使信念为真成为可能的。但高契尔的论证没有诉诸认识的理由。其主张将是：既然相信他的观点符合我们的利益，那么这一信念就是实践上合理的。当我们有实践的理由去致使自己持有某一信念时，这些理由并不支持该信念，因为在相关的方面，它们与该信念的真实性没有关联。

这一要点可表达如下。高契尔主张，相信“某些行动是合理的”符合我们的利益。他得出结论说，这样的行动是合理的。该论证假定

(D) 如果相信“某些行动是合理的”符合我们的利益，该信念就是真的。

然而，高契尔正当地拒绝了(D)。他设想有一个对各种关于合理性的信念做出酬答的恶魔。接着他声称，如果存在这样的恶魔，那么“持有关于合理性的错误信念就是合理的”。在此高契尔承认，尽管持有这些信念符合我们的利益，但它们仍然是错误的。它们符合我们的利益这一事实并不能使之成真。

高契尔能够收回这一主张转而诉诸(D)吗？看来他显然不能。假设高契尔的恶魔酬答这样的信念：我们的行为要是合理的，我们就必须名叫“贝蒂”，且戴着粉红蝴蝶结领带。高契尔不可能主张，如果存在这样的恶魔，该信念就是真的。我们也不需要离奇的情形去驳斥(D)。拥有关于合理性

的某一信念可能符合一些人的利益，而拥有相反的信念可能符合另一些人的利益。高契尔不可能主张，这些信念都是真的。

我们既然应该拒绝（D），那就应该拒绝支持高契尔观点的论证。即便相信高契尔的观点符合我们的利益，或者说使自己相信其观点是合理的，这也并不表明高契尔的观点是正确的。

这一论证可能揭示了某些东西。高契尔可能仍会主张，相信他的观点是实践上合理的。但除非主张其观点是正确的，否则他就不得不放弃其主要的目标。他无法坚持认为，道德地行动是合理的。他只能主张，这个信念是某种有用的幻象。

五

在对核威慑的讨论之中，高契尔给出了支持其观点的第二种论证。他认为，如果我们受攻击，那么产生报复的意图就可能是合理的。接着他主张，既然产生这一意图是合理的，那么如果威慑失败，依此意图行动就是合理的。

445

刘易斯（Lewis）拒绝这一推理。他虽然同意意图报复可能是合理的，但否认报复本身是合理的。

高契尔在其答复中否认“对于一个合理的策略来说必要的行为，其本身可能是不合理的”。如果接受威慑策略，他写道，那我们“拒绝它们所要求的行为就不可能是前后一致的”。既然“不可能主张这样的行为不应该执行”，我们就不可能称之为不合理的。“将某一行为评价为不合理的，就是……主张它不应该被……执行”。

这样的报复行为对于威慑策略来说不可能是必要的，因为如果该策略成功，这些行为就甚至不会被执行。但这是威慑的特殊性质，我们可置之不论。在我们所关注的大多数情况下，相关行为会被执行。因此，如果我由于该习性会符合我的利益而成为可信赖的，那我就必定预期我会守诺。与此类似，在先令情形中，我必定预期我受药物控制的状态会影响我的行为。在这两种情况下，如果我采取将有利于我的策略，我就必定预期会以将不利于我的方式行动。

接着要指出：即便在这些情况下，我的行为也不是为我的策略所要求的。对于我的策略之成功来说，它们是不必要的。如果必要，且我的策略有利于我，那我的行为就不可能对我不利。对我的策略来说必要的并非我的行为，而只是我的意图或习性。我的行为只不过是受欢迎的副效应。

我相信，这一区分动摇了高契尔对刘易斯的答复。如果某一策略尽管有负面效果但仍是合理的，那我们就可能同意，在某种意义上，这些效果“应该发生”。但这仅仅意味着，“事情应该如此这般地进行，乃至它们会发生”。并且就接受这一主张而言，我不必赞成或欢迎这样的效果。如果我们要举办宴会，那事情就应该如此，乃至我们随后不得不洗餐具。我们仍可以有理由对不得不洗餐具感到遗憾。类似的主张适用于产生于有利习性的行为。我们能够同意，在某种意义上这些行为应该被执行。事情应该如此乃至这些行为会被执行。但我们仍可以前后一致地认为，这些行为是令人遗憾和不合理的。

六

高契尔提出了支持其观点的又一论证。他声称，这一观点避免了合理利己主义的“某些不受欢迎的后果”。其中主要的后果是：按照那种理论，“成为合理的”可能是一种诅咒（curse）。

我认为，这一论证没有为高契尔的观点提供支持。高契尔承认，即使按照他的理论，成为认识上合理的也可能是一种诅咒。如果认识的不合理性（epistemic irrationality）直接得到回报，那么情况确实如此。高契尔宣称，任何理论都不可能避免这种不受欢迎的后果。但情况并非如此。高契尔可能将其观点予以引申。他可以做出类似的主张：我们的理论推理当且仅当符合我们的利益，才是认识上合理的。按照高契尔观点的这一版本，认识的合理性绝不可能成为诅咒。然而，这种修正并不会对高契尔的观点有所助益。在疯狂的推理符合我们的利益时，那并不会使之成为合理的。

如任何优秀的理论都会承认的，认识的不合理性可能符合我们的利益。实践的不合理性亦如此。这两种不合理性都可能得到回报。假定或接受这些事实的存在，这绝不会对合理利己主义构成反驳。

在支持其观点的过程中，高契尔还提出了另一种主张。他承认，其观点在运用于先令情形时看上去可能违反直觉。我们可能对如下主张迟疑不决：我的疯狂行动是合理的。但高契尔提出，这绝非反驳，因为“只要我们可能凭直觉倾向于认为……‘合理性’在帕菲特的质疑与批判之中都是某种技术性的用语”。

情况并非如此。我在问的是：在日常的意义上，想要什么和做什么才是合理的。而高契尔声称，先令情形“表明我们关于合理性的日常观念……有时是错误的”。高契尔既然在坚持认为应该修正我们的日常观念，那就无法为其“合理

的”的用法辩护——他是通过使之成为一种纯粹的规定来使用的，这按定义是正确的。而这也会使他的观点变得微不足道。

按照高契尔的观点，行为如果产生于有利的习性就是合理的。即便这样的行为仅仅是该习性令人遗憾的副效应，它们也是合理的；哪怕极尽疯狂之能事亦如此。这令人极其难以置信。我已讨论了在我看来高契尔支持其观点的一切论证。我的提议是，它们无一成功。我的结论是，我们应该拒绝他的观点。拥有某种习性可能符合我们的利益，使自己拥有它也是合理的，但依之行动则是不合理的。

高契尔提出契约论的一种霍布斯式版本，并为一种最低度的道德辩护，因为他认为由此就能说明：即便在自利的意义上，我们也被合理地要求绝不做不当之事。高契尔宣称，任何其他的道德理论都没有实现这一目标。如我所主张的，如果高契尔的论证失败，那么我们就失去了接受其最低度道德的主要理由。

附录C 道义式理由

448 在为规则后果论的康德主义论证之前提（E）辩护的过程中，我提出：

（X）如果最优原则要求某些我们认为是不当的行动，那么按照我们的观点，使该行动不当的特征或事实不会给予我们决定性的非道义式理由不这么做。真实的情况可能只是：由于使该行动为不当，这些事实会给予我们以决定性的道义式理由不这么做。

为了辩护（X），我们似乎可以诉诸如下主张：

（1）如果这些行动不是不当的，那我们就不会有决定性的理由不这么做。

但可能难以替该主张辩护。如果某些事实使得某些行动为不当，那就难以设想“这样的行动不是不当的”，因为可能不存在这一点为真的可能世界。而且即使我们可以诉诸（1），那也不会说明：正是这些行动的不当性给予我们决定性的理由不这么做。可能存在会使某些行动为不当的事实，当且仅当这些事实也给予我们决定性的非道义式理由不这么做。

我知道，对（X）的论证不会简短，此即我之所以仅提出（X）为真的原因。但有一个反对（X）的论证值得讨论。某些人宣称某一行动是不当的，此

时他们的意思是：我们有决定性的道德理由不这么做。这些人尽管诉诸道德的理由，但会否认存在任何道义式理由。按照这一观点

(2) 某一行动是不当的，此时这一事实是这样的二阶 (second-order) 事实：某些其他事实给予我们决定性的道德理由不这么做，并且“我们拥有这些理由”这个事实不会进一步给予我们某种独立的或非派生的理由不这样做。

这一主张与 (X) 相冲突，因为 (2) 蕴含着

(3) 如果最优原则要求某些不当的行动，那我们就会有决定性的非道义式理由不这么做。

我认为，我们大多数人不在这种决定性的道德理由 (decisive-moral-reason) 的意义上使用“不当的”。我们既然在别种意义上使用“不当的”，就可以有正当理由拒绝 (2)。而且我相信，正是在我们现在考察的情况之中，(2) 看来最讲不通。如果最优原则的确要求某种不当的行动，那么正是我们现在考察的这种行动，其不当性可以最为言之有理地称作进一步给予我们独立的理由不这么做。在某些这样的情况下，我们甚至可以宣称，这些行动的不当性会给予我们不这么做的唯一理由。例如，如果某一避孕方法是人为的，那么这一事实就其本身来考虑似乎不会给予我们不这么做的任何理由。

这个例子没有说明：如果如我们大多数人所认为的，某些避孕方法不是不当的，(2) 就是错误的。就问 (2) 是否正确而论，我们无法行之有效地考察显然不当的行动，并问：如果这样的行动不是不当的，那么真正的情况会是怎样的。如我所言，这种反事实 (counterfactual) 或许是不可能的，或至少是难以想象的。但考察某些人如何改变其道德观点，这或许是有益的。在描述观点改变的过程中，我要重新描述这些人的信念，由此让它们与其用于这些人实际考虑的微调版本，不如用于我的虚构情形。先设想

炸弹情形 (Bomb)：失控列车驶向隧道，将轧死隧道之中的五个人。你可以通过扔一枚炸弹到列车前而救那五个人。但我站在附近，由此炸弹也会炸死我。

许多人相信，这种行动是不当的。在考虑这样的情况之后，某些人接受

450 优先原则（Priority Principle）：不杀人的消极义务（negative duty）优先于救人的积极义务（positive duty）。

就解释这一原则而论，这些人会主张

（4）用同时会杀死其他某人的方式去救多个人，这是不当的。

接着回顾

隧道情形：你可以让一列无人驾驶的失控火车变向，转到另一轨道，由此会杀死我而使五人幸免于死。

这一虚构情形尽管几乎没有实践意义，但得到了许多讨论，因为在许多人看来，该情形似乎是优先原则的反例。当他们考虑隧道情形时，上述原则的若干支持者改变了看法。这些人不再相信（4）。按照他们的观点，你通过让列车变向去救那五人尽管也会杀死我，但你的行动是道德上可允许的。这些人由此假设

桥梁情形：你可以救五人之命，其唯一方法是使我跌倒于火车前，由此致我于死，但也使得火车停止。

这些人相信，这种行动会是不当的。他们的结论是：通过让某种威胁变向以便少杀人而救多人生命，尽管这不是不当的，但是，借助杀死其他某人来救这些人会是不当的。

根据（2），某一行动的不当性没有给予我们进一步的独立理由不这么做。有些人相信这是正确的，因为对于我们有决定性的道德理由不这么做这一主张，某一行动是不当的这一断言没有给它添加任何东西。如果这些主张正确，那么这么问就总是足够的：我们是否有这样的决定性理由不以某种方式行动。我们永无必要作为某种单独的问题去问：某一行动是否不当的。

如我所言，正是在我们现在讨论的几种情形之中，这些主张看上去最不

可行。我刚刚描述过：在比较像炸弹、隧道和桥梁之类的情形时，一些人如何改变其道德观点。这种观点改变不是关于我们以某些方式行动的理由之强度。在考虑隧道情形时，这些人不是先确定：你会有充足的理由让列车变向以救五个人；然后得出结论说：由于你有这样的理由，该行动就不是不当的。他们首先想到的是：救这五个人的方式不会是不当的；其中一些人进而断定：既然该行动不会是不当的，你将要救多人生命这一事实就给予你充足的理由而以这种道德上可允许的方式行动。类似的观点适用于桥梁情形。在考虑该情形时，这些人不是先确定：你会有决定性的理由不借助杀我来救那五人；而仅在此之后才断定：该行动会是不当的。这些人首先想到的是这一信念：该行动会是不当的；而仅在此之后才断定：该行动的不当性，给予你某种进一步的且或许是决定性的理由不这么做。

我已讲过：我们有些人是在某种不可定义的意义上使用“不当的”这个词，我将其含义表达为“绝不能做”（*mustn't-be-done*）。正是诸如隧道、炸弹和桥梁之类的情形之中，我们可能看来最有道理去相信，某些行动在此意义上是不当的。在隧道和桥梁情形中，你都可以通过以某种同时会杀死我的方式行动来救那五人。就我的观点来看，在桥梁情形中作为手段被杀，绝不比在隧道情形中作为副效应被杀更坏。但对于这两种类似的行动，许多人认为，只有作为手段的杀人，才具有作为绝不能做之事的显著性质。这样的行动是出局的（*out*），或不可允许的。而如果某一行为绝不能做，看来我们就能够有道理认为，这一事实给予我们进一步的独立理由不这么做。这些都是如下主张看来最不可行的情况：某一行动是不当的，此时这一事实并没有给予我们任何进一步的理由不这么做。

如我刚刚论证的，我们如果能够可证成地拒绝（2），那就能够拒绝这种反对（X）的论证。因此，我倾向于认为：在最优原则要求某些行动时，我们绝不会有决定性的非道义式理由不这么做。

译后记

面对着帕菲特的跨世纪巨作《论重要之事》卷一的译定稿，我终于可以敲下“译后记”三字了。不过，此时的我却是无法像以往那样完全轻松起来。原因很简单，《论重要之事》的卷二更加厚重，内容更加复杂，艰巨的任务还在后头。不过，乐观地看，总算在年前完成了卷一的翻译，可以稍事休息，至少可陪家人好好地过个年。尽管《论重要之事》近1400页的厚度与其学术贡献、学界影响是绝对相称的，但还是导致翻译不可能在短期内完成。因此，为了既保证翻译的质量，同时又让中国读者尽快见到译本，译者与出版社协商后，确定分三卷出版。卷一（前三部分）是帕菲特《论重要之事》的主要理论阐述部分，将于2015年出版；卷二的第四部分是四位作者对帕氏著作的评论，而第五部分是帕菲特对这些评论的回应，这二部分将组成卷二（上）于2016年出版；而卷二的最后一个部分，也就是有关元伦理学的第六部分，将作为卷二（下）于2017年出版。当然，之所以这样分开出版，也是由此分开的内容既有依存，但也具有相对的独立性，由此不会对我们的阅读造成太大的影响。

当初刘训练兄替出版社问我是否有意翻译这部巨著时，心里着实犹豫了好一阵子。因为一旦接下，则意味着未来几年要有大量的时间放在上面，所谓的寒暑假肯定是要泡汤。不过，最终有两个原因促使我鼓起了勇气，接下了这个艰巨的任务。第一个原因是我非常喜欢帕菲特的文风以及其论证的力度。最开始接触帕菲特的*Reasons and Persons*时，尽管知道它非常好，但由于我跨专业来学伦理学，既要花大量的时间来恶补哲学的基础知识，还要花大量时间了解分配正义的研究，因此没有多少时间认真地去阅读它。直到2008年进入华师

大的思勉人文高等研究院，有了大量可自由支配的时间后，我才能真正进入到该书的内在论证脉络。还记得当时特别兴奋，就像发现了新大陆一样，恰逢当时政治学系的江绪林兄也对帕菲特兴趣盎然，当时我们一旦碰到一起，聊的就是帕菲特。之后又从网上下载到了《论重要之事》的前身*Climbing Mountain*，再然后是下载到了《论重要之事》一直流传在外的草稿版，当时真是有如获至宝之感。尔后11年又极幸运地从刘擎教授那里得到了《论重要之事》的电子版。因此，翻译《论重要之事》之前，我对该书的内容已经非常熟悉了。而我阅读帕菲特所受的影响首先是对论证风格的认识。作为混在哲学界的理科生，语言上的修辞与文采一直是我的致命伤，而帕菲特顺承西季威克，试图把伦理学发展为一门科学，其中展现的正是典型的科学式理性思维，而这正是我的最爱。因此，帕菲特为人所讨厌的诸多思想实验，不停地追问各种逻辑可能性的论证，恰恰是为我打开了一扇窗，发现了哲学的另一种做法。另一个影响则是实质观点上的，接触帕菲特之前我勉强可以算是一个准后果论者，但对后果论受到的诸多典型反驳心有余悸，理论立场也颇为摇摆。而帕菲特的开创性思维，让我明白了许多看似强大的反驳实际上疑虑多多。正是帕菲特的*Reasons and Persons*对我的冲击，让我对后果论具有了崭新的认识，逐渐变成了一个坚定且激进的后果论者，估计比帕菲特本人都有过之而不及。而国内对后果论的研究由于对功利主义的误解，一直颇为薄弱，帕氏的这本巨著如果翻译过来，希望应该会有助于改变这种状况。

第二个原因则是好友阮航应允与我一起合译。即使是分三年来译，但近1400页的篇幅依然是让人望而生畏的。若找不到合适的合译者，肯定是不敢接手翻译的。阮航兄在湖北大学哲学系任教，我与其相识已久，相识的缘由正是起因于其翻译《织梦者》一书，正是在这过程中了解其对翻译特别“较真”，译文不仅忠实可信，文字更是简洁漂亮，而后者更是我所欠缺且羡慕的。为此还曾笑谈我们应该合译，我更多负责内容上的“信”，他更多负责文字上的“达”。当机会来临时，阮航兄尽管诸多事情缠身，但还是爽快地应允合译此书，承诺把此事作为头等事情，为其留出足够的时间。

确定接下时代华文书局的翻译后，阮航与我迅速商定了分工。阮航译第三部分和附录，以及与之相对应的内容提要、尾注和引注，葛四友译其余的部分。我们初步确定，用6个月左右的时间拿出各自完整的译稿；然后交给对方一字一句地校对，为时大概三个月；随后用一到两个月的时间各自针对对方的意见来修正自己的译稿，在此过程中商讨疑难句子的翻译，尤其是统一关键技术

语的译法，最后交由葛四友统一处理，形成定稿。在某种意义上讲，我们的翻译进度基本上符合我们的设想。不过，两人的翻译尽管都算认真，但校对时还是发现不仅各有错译之处，而且还有少量的漏译。幸得这种一字一句的严格校对，才让错漏之处一一显形。阮航兄尽管早前对于帕菲特并不是很熟悉，但一方面帕菲特的行文简洁明了，所言所论均是深入浅出，另一方面则是阮航兄有着出色的领悟能力，这使其翻译质量远超我预期。因此，负责校对翻译内容之“信”这一块的我，相对而言反而是轻松的。相反，阮航兄不仅指出了我的错漏之处，更是对我的行文进行了通盘的修饰，总量约三分之一的译文是其润色过的。对比我之前的翻译，读者就会发现此言不虚。结果自然就是，阮航对葛四友译文的校对付出的劳动更多，不过，葛四友在最后统一译名译法等各种杂事上付出的劳动更多。故总体上来说，两人的工作量是差不多的，而这次的翻译也体现了真正意义上的合作，没有主次之分。

尽管合译会存在着理解、译名选择以及译文风格上的差别，由此给读者的理解带来麻烦，但我们都认为，这次合译的效果显然优于独译，也超出我们最初的预期。通过彼此校对，译文质量大为改善，其利亦远远超过合译通常带来的不便。而对于译者就原文提出的一些疑虑以及注释中的一些页码问题，帕菲特教授在繁忙之中也做了详尽的回复，这自然也有助于译文质量的提升。当然，无论我们花了多少时间，做了多少努力，译事毕竟是无止境的，译文肯定存在诸多可改进之处，也期待方家能够提出宝贵意见，帮助我们进一步提高译文质量。

葛四友

2014年1月29日

于华东师大人文楼

有些注释是可以独立阅读的，因为我引用了足够的段落以用于这些注释。在其他注释中，我引用了文本的开头语以及随后的结束句。在后面的引注中我给出了引文。

xxxiv 西季威克的书尽管很长，有些章节现在也可以被忽略……我们可忽略卷1的第2章，卷2的第6章和卷3的第12章。但我们应该阅读第6版序言。

xxxvi 不过《方法》的第一稿就只有几个笑话，后来还删掉了。例如，西季威克去掉了这个主张：如果效用主义者知道，若共同体具有大量的美德，还具有些许“被人们共同责备的恶德”，则共同体的幸福可以得到增强，并且他们认为“其他人会提供这些美德”，他们就可以“认为自己提供这些恶德就是有证成的”（451）。由此义务的秩序……当问及是否为《方法》一书而自豪时，西季威克的回应是，“书的第一个词是伦理学，最后一个词是失败”。

xxxviii 西季威克其实是在进行颠覆之时……尽管有时候稍有冒险……西季威克写道：“如果我们考虑的是性与个人完善之间的关系问题，非常清楚的是：如果个人的性关系纯粹是感官之类的，他就会失去其情感性质最高和最好的发展，但我们很难先验地知道这类关系会影响到更高关系的发展（经验也并不表明情况普遍就是这个样子的）。后一种论证思路还有更深一层的困难。我们不得不为之辩护的公共舆论并不是纯粹通过与更高类型关系的比较来谴责较低类型关系的发展，而是根本没有任何比较对象。尽管我们并不谴责独身（若独身不是为了更高贵的目的，可能会鄙视他），但很难表明我们为什么应该谴责（仅就它与个人情感完善的关系而言）纯粹由性欲关系所提供的不完善发展”（ME 359）。威廉斯在提到西季威克对“贞操”一词的使用时，称此“无疑是布鲁兹伯里（Bloomsbury）觉得压抑和乏味的部分”（Williams [2003] 283）。这里有一个奇怪的颠倒。尽管摩尔也称“色欲的快感……享受它肯定本身是一种恶”（Moore [1903] 209），但布鲁兹伯里群体没有

觉得摩尔乏味。

xl 尽管西季威克也犯错，我在脚注中提到了一些……西季威克写过，“我不会为了将世界整合成一个体系而动一下手指头”，但他后来的确那样做了。西季威克写道：

“我们如果不以普遍幸福作为共同目的而把人类的行动系统化，那么还有什么其他的原则把它们系统化呢？”（ME 406）。他本不该假定：我们是要系统化这些行动，故我们应该是享乐主义者。西季威克拒绝了一切重要的分配正义原则，我认为这是错误的。他本该更清楚地区分这两个观念：“我们道德上应该做什么”与“我们最有理由做什么”。他做出了一些纯粹错误的主张，如他写道，“我认为，在现代文明社会，如果‘一个平凡的人’的良知被公平地引向对如下假想问题的思考，即若他追求一己幸福会给他人带来更大损失（没有任何其他抵消性收益），在这样的场合他这样做是否正当，那么他应该毫不犹豫地做出否定回答”（ME, 382）。

xlii 我假定永远不存在明显的错误，无论如何不会是重要的错误。罗尔斯很有风度地补上：“我总是理所当然地认为我们正讨论的这些作者要比我聪明得多。若非如此，我为什么要浪费我和学生的时间来学习他们呢？”（Rawls [2000] xvi-ii）。既然哲学能取得进步，我们现在就能看到许多比我们聪明的人所犯的明显错误。

xliii 斯多葛学派有关自杀的原则……，正如一个斯多葛主义者所说，“如果你的帐篷着火了，离开它”。

xliv 我们应该忽略这种情感的爆发。我们也应该忽略其他某些充满激情的伟大作者的情感爆发，诸如Ruskin有关Palladio Venetian教会的轻蔑评论。他称救世主教堂（Redentore）是一个破旧而下贱的乡下教堂。他在讨论圣乔治奥教堂（San Giorgio）时写道：无法想象，还有哪个教堂设计会比它观念上更粗俗、野蛮和幼稚，抄袭得更一板一眼，效果上更索然无味，在理性的每一种视角之下更可鄙（Rushkin [1903] 381）。

我像西季威克一样，有时候发现他“完全就是个启示”（HSM 151）。西季威克评论的是有关康德的术语。但他继续说道，“我们必须回到康德，从他那里再出发。我也并不打算自称康德主义者，而是始终要把他作为我的老师之一来看待”。

第31页 我们也许是整个宇宙中绝无仅有的理性存在者。如果我问，“我们是吗？”，无论作何回答都是令人震惊的。还有为数不多的几个这样的问题。

理由是由事实给予的。我们可以主张，理由不仅仅是由事实，而且也是由其他范畴（诸如心理状态或属性）中的事实所给予的。例如，有些人会说，欲望给予我们理由，行动的不当性给予我们理由不做它。但是所有理由都可以重新描述为由某些事实所给予的，诸如有关我们欲望的事实，有关行动不当性的事实。

第32页 ……我们最有理由要做的事情。当我们主张更有理由或最有理由以某种方式行动时，我们使用“理由”这个词不是指向具体理由的可数名词（如“树”和“湖”和“母牛”），而是用作像“木材”、“水”和“牛肉”这样的集合词，以指涉某个或某组理由而无需在上述具体理由之间做出区分。这同样适用于我们具有充分的理由和决定性的理由以某种方式行动。

第33页 ……“应该”或“应当”去做的。我认为，“应该或应当”所表述的概念如同理由概念一样，无法得到有益的定义。有些人表示，当我们主张我们应当做某事时，意思是我们有决定性的理由做某事。但这看来只是我们部分的意思或蕴含。

“应当”一词似乎加了点东西。其他人则试图通过应当来定义理由。我不相信这种定义能成功。但是即使这两个概念都是无法定义的，它们也有着紧密的关联，有些方式可以给出部分解释。我们可以部分地确定这个版本的“应当”，说这个概念适用于某个行动，正好当且因为我们具有决定性的理由或最有理由这样做。

第35页 如果我们说错误的信念能给予人们理由，我们需要加上这些理由并不具有任何规范力量。尽管我们最好是主张实践的理由只能由事实来给予，但是总体来说，可能还是承认“某些种类的错误信念也能给予人们认知理由”为好。

当我们相信自己具有某种理由时，我们具有了一类不同的显见理由。这种信念引出了特殊的问题，我会在第17节加以讨论。

第37页 ……就将鲜少讨论驱动性理由。我们从两个方面来看待这些理由是可接受的。基于心理学论说，驱动性理由是信念。基于非心理学论说，这些理由是我们信念的内容。我们的信念若为真，则非心理学论说是更自然的。如果人们问我为什么不吃核桃时，我会说“它们会杀死我”。但若我后来发现医生搞错了，我就会转而说，“我过去不吃核桃是因为相信它们会杀死我”。我们还可以把驱动性理由要么描述为我们想要实现的事物，要么描述为我们的目标或欲望。当人们问我为什么不吃核桃时，我可能说“为了活着”或“我想活着”。（正如我将论证的，第二个主张并不蕴含我的理由是由欲望所给出的。）

我认为，我们没有必要在驱动性理由的心理学论说和非心理学论说之间加以选择，因为它们是可兼容的，我们都能用。然而，两者的可接受性可能引起混淆。根据一种论说，驱动性理由是真实或显见的规范理由，是我们的如此信念在解释我们的决定和行动时相信的是什么。根据另一种论说，驱动性理由就是驱动性的心理状态。既然驱动性理由既可以看作是规范性理由，也可以看作是驱动性的状态，这可能表示规范性理由就是驱动性状态。这会是一个极大的错误。

第44页 如果我们将某事作为目的……我们的欲望是工具性的。当我们想要某种医疗检测，从而表明我们没有某种病时，我们具有的是其他类型的欲望。

第45页 对象给予的理由……我们也能称这些客观理由理论是基于价值的。如果正如我主张的，这些理由真正立基的不是这些结果或行动的价值，而是基于使得它们或好或坏的事实，那么这个词就有点令人误解。但有些客观主义者会拒绝这个主张。我使用“基于价值”这个词，在部分程度上是为了我的主张能涵盖其他版本的客观主义。

第52页 此种事件会是外在地好的，通过使得更长的事件系列成为……内在地好的。例如，基于某种观点，“罪有应得”是好的。由于“报应”取决于受惩罚者是否犯过某种罪，因此“报应”会是一种外在的属性。但按照这种观点，“某人犯罪随后受惩罚”这种长系列事件内在地没有“此人犯了罪但从来没有得到报应”这种事件

那么坏，从而前一事件在这个意义上是更好的。

第54页 ……如果我们不厌恶这种感知，那么无论是它们还是我们的意识状态都不是坏的。“痛”这个词是含糊的，它可以用来指称两类感知，要么（1）仅当它们是被厌恶的，要么（2）不管它们是否被人厌恶。心理学家为检测不同人的疼痛反应，缓慢地减低某种疼痛的强度，此时有些人会说，“再也不痛了”，有的会说，“还是有点痛，但我不再厌恶它了”。我只在第一种蕴含厌恶的意义上使用“痛”。这是具有规范重要性的那种意义。具有通常为人所厌恶的某种感知，但我们事实上并不厌恶它时，这并不是坏的（参见Kahane [2009]）。

第60页 然而，在有些情形下……主观主义者应当否认这种欲望给了我理由。还有一种描述这样情形的方式。主观主义者不是说我们没有任何理由满足这样的欲望或实现这样的目标，而是说我们不可能满足这样的欲望或实现这样的目标。如果你并不应得痛苦，我伤害你就不会给予你任何应得的痛苦，而如果你没有伤害我，我就没有什么可报复的。主观主义者可能会说，这种目的式欲望应该被看作是暗含了某种形式上的条件。我们真正想要的是：某些事实若确如我们认为的那样，则某件事发生。这些事实若不是我们认为的那样，这种欲望就不可能被满足，这可能就是它们为什么无法提供任何行动理由的原因。

第70页 我们可称一种规范性主张为实质性主张……。与这种意义的“实质性”相对的是“概念性的”或“无足轻重的”，而不是前面的“程序性的”。（我应该加上，有些概念性主张并不是无足轻重的。）

第88页 还有另一个理由……基于欲望的理由与欲望之链。或许有人提出，我们的欲望不是形成这样一种链条，而是形成一个循环，这样我们就有基于欲望的理由来具有所有这些欲望。但我们不可能有基于欲望的理由要具有这种循环的欲望。例如，假定我想要一些泥以便让碗里放点东西，而我又想要一只碗以便泥有地方可放。尽管可以主张，这两种欲望中的任一种都给予我某种给予欲望的理由去拥有另一种，但这对欲望并不给予我基于欲望的理由想要它们。我想要泥和碗并不给予我理由想要泥和碗。

我们现在可以回到情形2……这个情形提供了我们所称的、反对理由的主观主义的第二种痛苦论证。这个论证在某种程度上比我在11节给出的那个论证要强。那个论证指出，按照主观理论，痛苦的性质不给我们任何理由想要避免处于痛苦之中。这第二个论证指出，按照这些理论，我们可能具有决定性的理由想要我们为了痛苦的缘故而处于痛苦之中。这些理论的这种含义在某种意义上更难让人相信。然而，主观主义者可能主张，尽管他们的理论应用于虚构情形时具有讲不通的含义，但我们的欲望在大部分情形中给予我们理由。这至少在部分程度上回应了第二个论证。但是当主观主义者考虑第一个论证时，他们哑口无言。即使我们的欲望给予我们理由，它们也不能表明我们没有任何理由想要我们的欲望。主观主义者必须承认，按照他们的观点，被烧和被鞭打的感觉不能给予我们任何理由想要避免被烧和被鞭打。我们甚至没有一丁点的理由要相信这个主张。

第91页 主观主义者必须拒绝这些主张。这些理论还受到其他的反驳。假定那个反复无常的暴君威胁我，除非我明天中午具有为了折磨自身的缘故而有受折磨的目标，否则他就要折磨我。按照基于目标的主观理论，既然我现在具有不受折磨的目标，我现在就有基于目标的理由，要具有受折磨的目标来实现“不受折磨”这个目标。但是如果我成功地具有这个目标，这会给予我一个基于目标的理由要使自己具有不受折磨的目标，而这会给予我一个基于目标的理由要使自己具有受折磨的目标，如此以致无穷。这样我会在这些目标之间左右摇摆。这样在明天中午，我只有二分之一的机会具有使得我明天不受折磨的目标。基于客观理论，既然我服从专制者之威胁的理由不是由我的目标，而是被我具有和努力实现这个目标的不变理由所提供的，那就不存在这个问题。

第94页 然而，这些人要捍卫这个主张……然而，在某些情形中，这些人能够提出我在第60页的注释中所描述的那类主张。

第96页 尽管目的式欲望理论并不是不融贯的……我的其他反驳是适用于目的式欲望理论的。我已经论证，我们应该拒绝一切主观理论。我们接下来可以简要考虑一种混合理论。按照这种观点，我们要有理由努力满足某个欲望，就必须有基于价值的、对象给予的理由拥有这种欲望。我们想要的东西必须在某种意义上是好的或值得实现的。但是我们具有这个欲望若是合理的，我们具有这些欲望则会给予我们进一步的理由去努力满足这些欲望。当我们必须在同样好的各种可能目标之间做出选择时，我们的欲望或偏好就能打破这种僵局，给予我们理由采用其中的一个目标。我相信——尽管这一信念并非很强——我们甚至应该拒绝这种混合理论。我认为，这两种理论是水火不容的。当我们具有某种欲望时，这个事实可能使此为真：我们有进一步的理由满足这些欲望。但这些进一步的理由不是由“我们会满足这个欲望”这个事实提供的，而是由“依赖于我们具有这个欲望”的其他各种事实所提供的。我在第9节的末尾描述了一些这样的事实。尽管我相信我们应该拒绝这种混合的理论，但我反对纯粹主观理论的论证并没有表明这一点。这个问题仍将是开放的。但我认为，既然这种理论在根本上是客观主义的和基于价值的，这个问题就没有太大的重要性。

第100页 就法兰克福的真实观点来说……没有任何目的就其自身是好的。这个区分常常被人们忽略。在萨特有名的例子中，一个法国青年必须在照看母亲与加入抵抗队伍之间做出抉择。这两个选项都是好的，尽管没有任何一个明显更好。由此这个例子并不如萨特所相信的那样，支持主观主义的萨特式存在主义版本。

第101页 按照理由的主观理论……当任何孩子的生命得救时，还有很多人的知情欲望并没有得到更好的满足。如这段评论所蕴含的，“最好的”这种蕴含不偏不倚理由的含义，丝毫无关于对不管是结果善性还是道德的某种“不偏不倚的观察者”论说。这些论说把“什么是最好的，或正当的”定义为任何不偏不倚的观察者事实上会选择或赞成的。这种论说所获无几。我们若主张这种观察者只是具有不偏不倚的视角，就不能假定他们都会做出相同的选择。我们如果补充某些心理学假设，那么

也许能搞清楚这些观察者会选择什么。例如，如果这样的观察者是仁爱的，他们会选择有益于人们的事情。但这种预测没有任何重要性。

第110页 但我认为，按照这些观点，不可能存在任何真正的规范性理由，只存在行为的原因。因为这些作者认为规范性理由就是某种行动的原因，所以他们可能会拒绝这这种描述。很难用某种中立的方式来描述还原论的观点。

第111页 接下来我们可以更仔细地查看……我的信念所引起的，而且是由它所证成的。这些欲望要被这些信念所证成，就必须还是由这些信念以正确的方式引起的。我们在此没必要探讨这里涉及的是何种因果性。

第113页 当且因为某个欲望本身是合理的时候，我们具有这种欲望就在某种意义上是合理的。我这里区分的是“某种欲望或某类欲望本身”与“某人具有这个欲望”。如果你和我都希望威尼斯不被海洋淹没，我们具有的就是相同的欲望，但你具有这个欲望与我具有这个欲望不是一回事。在这个例子中，我们想要的是相同的事件。当我们想要的是不同的事件时，我们仍可能拥有在较宽泛的意义上相同的欲望。例如，如果我们在下棋，我们都想赢，情况就是如此。“某个信念本身”与“某人具有该信念”也具有类似的差别。因为欲望和信念既可指称欲望或信念自身，也可指某人具有这个欲望或信念，所以这两个语词都是含糊的。尽管我有时候会说我指的是哪一个，但人们通常还是会忽略这个区分。

第121页 ……所以他的偏好没有那么不合理。这种观点并非讲不通的，因为有其他的事实给予我们理由不那么关注那些更为遥远的事情，由此具有这种折扣率。比如这样的事实：我们对于遥远之事的信念，其正确的可能性通常较小；我们努力来产生或阻止遥远事件的紧迫性也更小；作为现在的我们与作为遥远未来的我们之间具有的心理关联很可能也是更少的；并且早期利益往往会产生其他的利益。但这些事实并不能证成一个纯粹的时间偏好，或者说一种就时间本身所打的折扣。（我在RP中从63节到67节讨论了这个问题。）

第128页 “内在地优于”这种关系是传递性的……。如果情况是，若A与B具有某种关系，且B与C具有这种关系，则A与C一定也具有这种关系，那么这种关系就是传递性的。例如，“是某某的父母亲”这种关系就不同于“是某某的祖先”这种关系，后者是传递性的。

第132页 下面两种情况所提供的理由就不是类似的……可能并不存在精确的事实。这里值得提一提有关快乐或痛苦之强度的各种主张是含糊的。“强度”的一种意义是纯粹心理学的。例如，当人们使用这种意义时，他们可能主张

(A) 与具有一小时这种强度的痛苦相比，具有十小时一半强度的痛苦会是较好的。然而我们通常假定

(B) 有些苦难的痛苦性质直接取决于其痛苦的强度与持续时间。例如，如果我们的痛苦强度加倍，或者痛苦的时间加倍，我们的痛苦就加倍。

如果我们假定(B)，(A)就蕴含着，具有更长时间的苦难会是较好的，尽管较长时间的苦难会是更为痛苦的。基于这种观点，情况就常常是，具有更多的痛苦本身

会是更好的。这些结论是说不通的。我们最好是这样主张

(C) 如果两个苦难之一会是更痛苦的，这个苦难本身就是更糟的。

要能够同时主张(B)和(C)，我们就必须在第二种意义即评价意义上使用“强度”。例如，我们应该主张

(D) 如果具有一小时的强烈痛苦与具有两小时不那么强烈的痛苦是一样坏的，这个不那么强烈的痛苦就应该只有前者一半那么强烈。

我们的主张不应该是(A)，而应是

(E) 如果十小时的痛苦比不上一小时更强烈的痛苦，那么第二种痛苦应该是比第一种强十倍。

正如这些评论所表明的，尽管我们有时候需要在心理学意义上来使用“强度”，但评价意义上的强度才是更重要的。许多作者忽略了这个区分，因此做出了错误的主张。这点也更明显地适用于快乐。正如西季威克所写的，“我们必须要小心，不要把快乐的强度与感知的强度搞混了，正如快乐的情感可能是强烈且引人入胜的，然而可能不如另一种更为精美和微妙的情感。

第133页 ……在部分程度上诉诸了个人同一性的合理重要性。既然西季威克认为人的分立是根本且极具规范重要性的，他这么不重视分配正义原则就是令人吃惊的(ME 416-7)。

第135页 按照西季威克的观点，我们能够理性地做这样的事情……西季威克并没有考虑这样的情形。他如果已经考虑到，也许会对其观点做出限定。

第137页 广义的基于价值的客观观点：一个行动使得……类似的主张也适用于其他情形。例如，假定我们能够或者(1)使某个陌生人免于十小时的痛苦，或者(2)使自己免于2小时的痛苦，或者(3)使这个陌生人免于五小时的痛苦和使自己免于1小时的痛苦。尽管(3)既不是不偏不倚地最好的，也不是对我们自己最好的，广义的基于价值的客观观点可能蕴含着，作为一种折中，我们应该理性地做(3)。

第142页 义务与自我利益的二元论……按照我早前所称的西季威克的实践理性的二元论，我们能够理性地做不偏不倚意义上最好的行动，或者做对我们自己最好的行动。西季威克并没有区分二元论的两个版本，因为他认为我们的义务总是做不偏不倚地最好的行动。西季威克对其二元论的其他评语引起了有趣的问题，我希望在其他地方做出讨论。

第143页 然而，这些令人印象深刻的严峻主张……西季威克还写道：“当个人富有激情地拒绝同意‘德性的报酬’是‘尘埃’时，通常情况下其出发点与其说是对自己报酬的私人算计，不如说是公正地避免一个在根本上如此不理性的宇宙，其中“个人的利益”最终并不等同于“普遍利益”。(ME 第一版，471-2)

第151页 在证据相对的意义上是正当的……这种意义可以具有两种版本，一种指称我们实际上知道的证据，另一种指称在我们能让自己意识到的意义上可用的证据。事实相对的意义通常被称为“客观”的意义，但是这个词还有其他的意思，其他的意思并不适合称为“主观的”。还有一种道德信念相对的意义，这是我们马上要接

触的。

第157页 按照康德式观点……我在注释中提到了一些这样的方式。首先假定某个人正非常小心地驾车，但有个小孩子冲到车前被轧死了。既然这个人完全没有疏忽，他就不应受任何责备。但是正如内格尔（1979）指出的，我们能可行地认为，这个人应该为他轧死了这个孩子的事实而感到惊骇，即使他是完全无可指责的。威廉斯（1981，27）称此反应是“行动者遗憾”。如果这个人由于知道他无可责备而高兴地吹着口哨驶离，那就是令人心寒的。（我们可以顺便提及，“行动者遗憾”这个术语看来过于狭隘了。假定还有个人因为某个不可见的障碍而无可责备地被绊倒了，倒下来时压死了一个小孩。我们会指望这个人对于引起这个孩子的死亡怀有类似的遗憾之情，尽管这个死亡不是任何行动的结果。）

接下来考虑两个同样不小心驾驶的人，其中一人杀死了一个小孩子。这两个人是同样可责备的，都应该感到悔恨。许多人会认为，杀了人的那个人是更可责备的，也应有更多的悔恨。但有此信念的那些人可能在部分程度上假定：这个人也应该体验到行动者遗憾，对其作为的感受要糟糕得多。另一个没有杀人的人就没有任何理由具有这种行动者遗憾。不那么清楚的是，杀死孩子的这个人除了感觉到这种行动者遗憾外，是否还要感觉到更大的悔恨。

我们接下来考虑边沁的可行原则：所有的惩罚都是坏的，由此在效果相同的前提下，我们应该尽可能少地惩罚人。未遂的谋杀具有一种不同寻常的犯罪特征：这是没有任何人意图去犯的一种罪。除非我们有意杀死某人，否则我们不会谋杀未遂。未遂的谋杀是未遂的谋杀。这样，如果对于未遂的谋杀不做任何惩罚，这不会导致任何人去犯这样的罪。边沁可能论证说，我们不应该有任何这样的惩罚，因为这种惩罚不会威慑任何罪行，因此没有任何好处。

人们可能反驳道，尽管企图谋杀者不会有意失败，但其中有些人认为他们有相当的可能会失败。惩罚未遂的谋杀可能具有一些好效果，即威慑某些这样的人。尽管如此，我们还有另一种理由对未遂谋杀的惩罚轻于成功的谋杀。如果惩罚是相同的，这可能威慑某些第一次企图谋杀者。但是，人们如果进行过不成功的企图谋杀，那就更有理由继续努力，直到成功。这种成功不会增加他们的惩罚（如果他们被抓到并定罪的话），并且通常可能会降低他们被抓的可能性。我们总体来说具有决定性的威慑理由，不像惩罚成功的谋杀那么严重地惩罚谋杀的企图。

我们可以补充说，在英格兰法律中，尽管大部分未遂谋杀是罪行，但是以不大可能成功的方法所做的未遂谋杀则是无罪的。类似的论证可以解释这样做的原因。我们都知道，如果这种企图不受到惩罚，也没有人倾向于这样去做。没有人会使用他们认为不大可能成功的方法，这样的企图也不会产生任何危害，但企图谋杀者在道德上可能是同样可责备的。有些人也许会认为，用针扎其所恨之人的蜡像，是在使用最有可能成功的方法。

我们现在回到这样的情形，两个人企图谋杀，只有一个成功，他们俩之间具有何种道德差别。尽管这两个人都应该感到悔恨，但成功的谋杀者应该感觉一种更大的悔

恨，这是谋杀未遂者没有理由要这样感觉的。出于威慑的理由，成功的谋杀者应该受到更严重的惩罚。这两个主张可能充分地描述了这两人之间的道德差别。这里可疑的是，成功的谋杀者更可责备，且如果我们相信应得，这个人应得更多的惩罚。正如我所说，这个观点是悖论性的。既然我们认为可责备性和应得的理由不能完全地取决于运气，不清楚的是，这些事情如何能够部分地取决于运气。

第160页 期望主义……期望上最好的。许多人谈论的不是这些结果的期望（expectable）善性，而是其被期待的（expected）善性。但这个词是误导性的，既然这种期望善性通常不是或应该被期待的善性。

我们同样能主张……行动效果的期望善性就大略是……期望主义并不需要假定，结果的期望善性只取决于可期望的利益总量。例如，正如布罗姆和卡恩表示的，这些利益或获得这些利益的机会在不同的人之间是如何分配的，这也是重要的。在涉及挽救许多人的情形中，如果每个人得到平等的获救机会，那么即使由此可能使得获救的人稍少一点，这也可能是更好的。我们也有理由是风险规避的，给予避免最差的结果以更大的权重。参见“Fairness”，载于Broome（1999）和Kamm（1993），第7章。

第166页 但我们最好不要把理性利己主义看作一种道德观点，而是看作道德的外在对手。西季威克尽管称理性利己主义是“伦理学方法”之一，却是在一种广义使用“伦理学”，后者涵盖了我们有理由去做事情的一切主张。

第169页 （K）我们应当不偏不倚地去做的……这些效用主义者不是主张我们应当要最大化幸福总量，而是主张我们应当最小化（去掉幸福之后的）痛苦总量。正如最小化净成本等于最大化净收益一样，这些行动方式也是一样的。效用主义者告知我们要最小化痛苦的净量，提醒我们如何最有效地使得有感知的存在者过得更好。其观点的这种表述更好地表达了“什么使得它是可行的”。按照这种观点的佛教版本，两种伟大的德性就是洞识与慈悲。

第186页 怀特不会具有充分的理由放弃自己的生命让我来救格蕾的腿。怀特若是老人，而格蕾是个年轻的职业舞蹈家，事情就可能不同了。怀特的损失可能不比格蕾的大。这类事实是道德上相关且是进一步的事实，而在考虑我的例子时，我们假定并不存在这种事实。

第209页 同意原则若应用于这两个选项……当然，还有其他的选项。这个人会有充分的理由同意我用钱让其他人避免同样大的伤害。同意原则于是会要求我做出这种捐赠。

第212页 纯粹手段原则……有些作者论证，我们可以忽略康德的这个主张：我们永远不应该把人纯粹作为手段来对待，因为知道康德“要把人作为目的来对待”的意指就足够了。这些作者主张，我们若把某人作为目的来对待就不会把他纯粹当作手段来对待。康德主张，把人作为目的来对待在部分程度上就是不把他们纯粹当作手段来对待。我们由此应该问这个主张蕴含什么。

第215页 既然相关性和重要性都是程度上的问题，（1）是否为真通常就是不清楚

的。正如这个评论所显示的，我提出（B）并不是作为一个准则，从而有助于我们决定某个人是否被纯粹作为手段来对待，或接近于这样做。（B）纯粹是描述我们如何可行地拒绝该行动属于这一类。没有任何人应该反对（B），从而主张，即使（3）我们对某人的对待在很大程度上受到相关信念或关注的管辖，情况仍然可能是（4）我们是在把这个人纯粹作为手段来对待。若（4）为真，（3）就不会为真，因为我们对这个人的对待不会在重要的或相关的意义上受到管辖。

第229页 有害手段原则。这个原则需要受到拓展或修正，理由是由Kamm（2007）在第4和5章所给出的。

第241页 基于康德的观点，和拥有善良意志一样……这个结论是疯狂的。尽管罗斯是个多元主义的直觉主义者，他并没有承诺某种僵化理论，他接近于具有类似的结论。罗斯考虑过但并没有坚定地拒绝这个观点：知识上的一丁点增加要比任何数量的快乐（或如他的其他主张所蕴含的，不具有任何数量的痛苦）具有的价值都要高。如此好的一个哲学家并没有拒绝这样一种可怕的观点，实在让人心碎。但这是在他早期。

第242页 “人性”这个词……康德是“伟大思想家中最不精确的一位……”（Kemp Smith [1962] xx）。作为一个例子，我们能够注意到康德误述了其观点。康德主张，人性是个自在目的，它具有在最高和无条件意义上的尊严。但康德还主张只有善良意志才具有这种最高价值。科丝嘉主张，这些主张并不冲突，因为康德使用“人性”来指称“理性选择的能力”，这种只在那些具有善良意志者那里才能得到充分实现，因为这些人的选择才是充分合理的。（Korsgaard [1996] 123-4）。这个建议在某种意义上是可行的。但是康德还用“人性”指称理性存在者，他主张它是具有最高价值的自在目的。我们无法言之成理地主张，理性存在者与善良意志者是一样的。我们也不能主张：这种存在者是与目的王国或最伟大的善（具有普遍德性和应得幸福的世界）是一样的。尽管康德主张只有善良意志具有尊严，但我们应该承认：按照康德的观点，存在几种事物具有这种最高的或无可超越的价值。（参见 Dean [2006]）

第247页 摩尔主张，当我们说……这个行动是正当的……由此产生最多的善。以摩尔的话说，“正当的”……确实且只能意指“好结果的原因”（Moore [1903] 196）。摩尔必定意指“最好结果的原因”。摩尔一般会加上，“重要的是坚持，这个根本点之肯定无疑是可证实的”。几十年来，摩尔的乌云遮住了来自西季威克的阳光，这在部分程度上是因为与审慎的西季威克不同，摩尔如同康德一样，其写作采用了那种极有说服力的极端主义。除了“有机统一体学说”之外，摩尔《伦理学》中的有趣主张，要么是来自西季威克，要么正如刚才引用所表明的，是明显错误的。正如威廉斯所写的，“摩尔哲学的特点是做作的谦逊谨慎，这一特性使得其行文拖沓，却没有阻止其犯下最狂妄的错误。（Williams [1985] 16）。（这些评论属于摩尔式的夸张）。

摩尔的《原理》并没有提出一个实质性的道德观点。令人吃惊的是摩尔竟然犯了这

个错误，因为他用了一整章来谴责这种错误，他称之为“自然主义谬误”（尽管它既不是自然主义的，也不是谬误）。西季威克用两句话就更准确地描述了这种错误。（ME 26 注1，和109）。

第275页 不可能性公式……G^①424。康德还写道，“有些行为的构成是这样的：其准则甚至一旦被视为普遍法则就会陷入矛盾之中……”有些作者追随奥尼尔将该公式称作“观念中的矛盾检验”（contradiction-in-conception test）。在已然确定这一点——在康德意指的意义上，某一准则要作为普遍法则会是怎样的——之时，我们可能发现：去设想某些准则是这样的法则，这会是逻辑上不可能的，且在此意义上是一个矛盾。但康德经常诉诸仅仅是经验的不可能性，因而我们也应该这么做。……（A）不会是一种有用的主张。如下情况也是不清楚的：在何种意义上，我们全都被允许奉行某一准则，这一点不可能为真。如果这样的行动是不当的，这就不可能为真；但这样无助于我们确定这样的行动是否不当。康德还主张，奉行任何这样的准则是不当的：我们不可能意愿“我们全都被允许奉行该准则”为真。这是一种更为言之有理的主张，我后面还会回到这个主张。

第279页 ……受道德激励的、包含信任的许诺这种社会实践也不再存在。可以补充说：我们如果全都相信这样的许假诺是可允许的，那就甚至不会理解某种道德的、包含信任的许诺这一概念。（可能仍存在某种类似于许诺实践的行动，但不同的是它采取某种非道德的形式。这样的许诺如同威胁。为了保持我们作为威胁执行者的名声，我们可以有理由执行威胁；恰与此一般，为了保持我们作为守诺者的名声，我们可以有理由遵守这样的诺言。）

第281页 如此种种情况表明，（G）是不可接受的。可能有人会说，这些虚构情形是不现实的，因为现实世界中的事实不会像我要求我们去设想的那样简单。但这些情形是可行的，足以对（G）的可接受性提供有效的检验。在运用于这种多少有些简化的想象情形时，道德原则应当是成功的。而且关于其想象的许假诺，以及关于若干其他的准则，康德的主张都同样地是简化的。

第286页 这两个公式既然相似……如我的一个注释所解释的……假设我们仅诉诸可允许性公式。我们由此就会问，我们能否合理地意愿如下情况为真：每个人都在道德上被允许奉行某一准则，即便这对于任何人的道德信念或行动都毫无影响。这不会是一种有益的问题。首先，难以设想的是，我们能够意愿“某些行动是道德上可允许的”为真。就大多数不当的行动而言，如康德以及许多其他理论家所声称的，甚至连上帝都不能意愿“这些行动不是不当的”为真。而如果某些行动是道德上可允许的这一事实对于任何人相信或做什么毫无影响，那么不清楚的是，除了“这些行动确实是被允许的”这个事实之外，我们还有什么理由意愿这些行动是被允许的。但这样一来，可允许性公式就无法帮助我们判别，这些行动是否是可允许的。

① G为*The Groundwork of the Metaphysics of Morals*（《道德形而上学基础》）的代码。

第289页 假设我做了件错事……康德的公式将错误地允许这样的不当行动。在此我认为，不同于康德的同意原则，康德的普遍法则公式企图成为我们所需要的唯一道德原则，乃至在康德公式的两个版本都未能谴责这一行动时，普遍法则公式蕴含着，该行动是道德上可允许的。

第292页 我们可以转而考察不同类的例子……他也不可能合理地意愿“每个人都相信那些救命的撒谎是不当的”为真。如我在第235页提出的，如果康德接受整体框架观（Whole Scheme View），那么对他来说，永远无人撒谎就可能并非不合理的。但是，整体规划观是错误的，而且在运用康德的公式时，我们应该问：如果了解相关的非道德事实且绝无错误信念，人们就能够合理地意愿什么。

第298页 在不清楚……但我们在此不必答复这些问题……在我们已经以其他方式修正康德公式的情况下，这种问题的重要性也较小。

第301页 在运用……按照康德公式的最佳版本——这似乎也是康德心中所想……如康德对自立者——他持有“不助人也不骗人”（G423）的准则——的一段评论之所示。康德主张：如果每个人都奉行这个人的准则，那么这样的世界将优于现实世界，现实世界中有些人助人但许多人骗人。但康德也主张：我们不可能合理地意愿“每个人都奉行这个人的准则”为真。因此，康德隐含的比较对象不可能是现实世界，而似乎是一个无人奉行这种准则的世界。

第302页 我们可将这样的情况称作群己困境。在RP^①第一部分及帕菲特（1986），我讨论过这样的情况。

第303页 ……我们就是在做肯定对我们所有人来说都较不利之事。也存在许多或然性的（probabilistic）群己困境，它们诉诸不同行动的可能效果，或者是对于人们来说预期最好的效果会是什么。

这样的主张并非关于……重复的囚徒困境。在最简单的此类情况之中：（1）我们每一个人通常可以要么自利，要么给予他人以较大的利益；（2）由于涉及的人数相当少，每一个人所做之事可能影响到在随后的情况下他人怎么做。例如，在两人的情况下，如果我给予你那份大利，你就可能回报给我大利。如果接下来我自利而获小利，你就可能报复以自利而获小利，从而不利于我。它们尽管被称作“重复的囚徒困境”，但不涉及哪怕一种真正的囚徒困境，或群己困境。在此类情况下，真正的情形并不是：如果每一个人而不是无人做肯定对自己更有利之事，那就对我们所有人对更不利。理论上说，这些情况要乏味得多，且较少根本意义，因为它们只是如下的多种情况之一：其中哪种行动会最有利于我们自己，这是不清楚的。这些情况实践上的重要性也低得多，因为它们要少见得多。然而，它们对于进化心理学家和历史学家来说很重要，前者力图解释动物行为与人类心理的各种特征，后者探讨在较早的世纪里，大多数人生活其中的小社群。

① RP为帕菲特著作*Reason and Person*（《理与人》）的代码。

群己困境尽管常常被忽视……而众多的情况都涉及众多的人。值得提及的是另一类情况，我们可称之为助人者困境（Samaritan's dilemmas）。我们每一个人都时而会通过自己付出较小但确实的代价或负担，去帮助某个有需要的陌生人。例如，如下情况的确可能发生：我们能够帮助某位遭遇不测者，或者是我们能够归还某个具有重要私人意义的失物。如果我们所有人总是给陌生人提供这样的帮助，那么与从不曾有人给予这样的帮助相比，这对我们所有人都更有利。但如果像世界上半以上的人的实际情况一样，我们住在大城市，那么对于每一个人来说如果不给予这样的帮助可能又会更好。这个人会避免自己付出成本。而他是否得到这样的帮助，极少取决于他是否给过他人此类帮助。我们每一个人未能帮助的陌生人，极少会是以后可能会帮助我们的那个人。因此，我们没能帮助他人，几乎从不曾导致他人出于不满而拒绝帮助我们。但是，如果无人帮助他人，那么尽管我们每一个人都是在做更有利于自己之事，但我们却是在对我们所有人更不利之事。

它们涉及公共善……可进一步区分出如下两者：一类公共善是事实上有益于那些即使并未出力的人；另一类则是必然会有益于这样的人，因为不存在任何可行的办法防止不出力者获益。洁净水通常属于第一类，清洁的空气属于第二类。

第306页 康德的公式不仅冲突于……这样的原则……在集体的层次是直接地自我挫败的。还存在另一种方式显示：在这样的情况下，常识道德可能其本身就蕴含着，我们应当不再优先考虑与我们M相关者。如果我们与相关群体的其他成员都可以相互沟通，并且我们相互之间都了解对方是可信赖的，那么达成某种联合的（joint）有条件协议，即我们总是通过给予他人以较大利益而做出有别的行动，就是对我们的合理的道德要求。如果仅当每个人都做出这样的承诺，这种联合协议才具有约束力，那么这一事实在我们决定是否承诺时就将我们的行动联结在一起。就做出如此承诺而论，我们每一个人都是在做最有利于自己或与己M相关者之事，因为他是在促进如下情况的发生：每个人而不是无人做较有利于自己或与己M相关者之事。由于这种协议要求一致同意，每一个人就都会知道：如果他不做承诺，那么整个规划就会失败。这就是我们的行动联结在一起的方式。因此，常识道德本身就要求我们既做出又信守上述承诺。然而，这种解决方案很少能实现，因为我们并非都是可信赖的，且即便都可信赖，通常也难以安排和实现这样的一种联合的有条件协议。我们如果都是有足够良知的康德主义者，那就会避免这种问题。（更充分的讨论参见RP第100-108页。）

第307页 康德的意思必定是……“做总体上会最大地促进每个人的幸福的任何事”的准则。然而，这种解决方案可能以另一种方式成为间接地集体自我挫败的。参见第404-405页及其后，以及RP第10-42节。

在大多数此前提及的贡献者困境中，上述主张都是正确的……以及我们孩子的孩子施加这样的伤害。而我们在此可以做个区分。显然，在群己困境中，我们全都应该采取的理想行动是给予他人以较大利益。如果与无人这么做相比，我们宁愿所有人都这么做，那对每个人都较有利。但康德的公式要求即使其他大多数人这么做

的情况下也要这么做。在此情况下，由于按此方式行动，我们将放弃自己可以获得的那份较小利益，而同时又不能从他人那里得到那份较大利益。这样的要求可能有时就变得极其苛刻，而且可能也是不公平的。例如在未解决的父母困境中，如果我们给予他人的孩子以那份较大利益而他人却未将这样的大利给予我们的孩子，那对我们的孩子来说可能就是不公平的。在至少某些这样的情况下，我们可以正当地认为：其他有多少人做我们在理想意义上应做之事，这会造成道德上的差别。我们可能被要求给予他人以那份大利，仅当有足够的他人在这么做。在其他情况下，作为防御性的次优选择，我们可能被允许将那份小利给予自己、我们的孩子，或与我们的M相关者。（关于怎样会是足够的一个提议，见RP第100-101页。）

第311页 我们依然没有完全答复……他没有孩子就是可允许的。奇怪的是，如康德的一位传记作家所言：“康德构想出‘绝不结婚’这一准则。实际上，只要康德想说明准则的某种极罕见例外是可接受的，他就会说，规则在此：‘不应结婚！但是，让这对相配的人成为例外吧。’”（见库恩 [Kuehn] [2001]，第169页。）

第319页 某些规则可能采取此类有条件的形式……类似的主张将适用于那些诉诸“如果人们接受某些规则那会怎样”的RC版本。我所提议的修改，更容易用于规则后果论的这些接受版，因为最优规则将采取简单得多的形式。（如里奇 [Ridge] 所指出的，即便这些规则采取有条件的形式，可能也不存在这样的一组规则：在每一种接受层次上其被接受都会使事情进展最好。但会有若干组这样的规则：在各种层次上其被接受将平均地或总体上使事情进展最好。里奇 [2006]）

第320页 ……能够言之有理认为，每个人都应当持有同样的道德信念。例如，如果人们持有相冲突的信念，那么，这些信念就不可能都为真；且我们还能认为，每个人都应当持有或者是努力持有真实的道德信念……简单地答以“大多数人不会”是不够的。另见RP第一章。

第321页 这条规则所表达的……至少三种最早的世界文明。它们是古代中东、印度与中国。一些引文如下：

佛教 (Buddhism)：勿以会伤害自己的方式去伤害别人（《优陀那品》第5章第18节 [Udana-Varga 5:18]）。

基督教 (Christianity)：无论何事，你们愿意人怎样待你们，你们也要怎样待人；这就是律法和先知的道理。（《马太福音》第7章第12段 [Matthew 7:12]）

儒家 (Confucianism)：己所不欲，勿施于人。在邦无怨，在家无怨。（《论语》第15章第23节 [Analects 15:23]）子贡问曰：“有一言而可以终身行之者乎？”子曰：“其‘恕’乎。己所不欲，勿施于人。”（《中庸》 [Doctrine of the Mean 13.3]）^①

① 出处有误，但此处原文如此。前一段引文出自《论语》第12章第2节。第二段引文出自《论语》第15章第23节。

印度教 (Hinduism)：这就是整修的律法；你自己不想经受的事，不要对别人做。
(《摩诃婆罗多》第5卷第1517 [Mahabharata 5:1517])

伊斯兰教 (Islam)：你若不为自己的兄弟渴望他为自己而渴望的东西，就不是真正的信徒(《圣训集》[Sunnah])。(《伊玛目瑙威四十段》之第十三 [Number 13 of Imam Al-Nawawi's Forty Hadiths])

耆那教 (Jainism)：人自己想受到怎样的对待，就应该怎样对待万物。(《苏特拉克里-坦加》[Sutrakritanga] 1.11.)

犹太教 (Judaism)：你不愿施诸自己的，就不要施诸别人。这就是整个律法；别的一切都是评注。(《塔木德·安息日》[Talmud, Shabbat] 31a)

神道教 (Shinto)：“你面前的人心如镜，从中可见你自身之形”。(黑住宗忠 [Munetada Kurozumi])

道教 (Daoism)：见人之得，如己之得。见人之失，如己之失。(《太上感应篇》[Tai Shang Kan Yin P'ien])

琐罗亚斯德教 (Zoroastrianism)：对己不利者勿施于人，其性善。(Dadistan-i-dinik, 94:5)

第327页 “每个人”一词……我们应当以该规则所要求的不偏不倚方式来考虑人们的福祉，而我们包含于这些人之中。康德同样写道：“既然将我自己排除在外的所有他人不成其为所有人，那么仅限于此的准则就不具有法则的普遍性……使仁慈成为一种义务的法则将包括我自己于实践理性的指令范围之内，以作为仁慈的一个对象。”(MM^①第450页)

第338页 在康德关于其公式的所有说法之中，没有任何一种会支持内格尔的解读。康德的确写道，“每个理性的存在者……必定总是既从自身的角度也同样地从每个理性存在者的角度，来看待他的准则”(G第438页)。但这一评论并非来自对其普遍法则公式的讨论，而是来自对目的王国公式的探讨。而如果康德原本有这样的意图，即我们应该去设想人之待我如我之待人，那他就不会如此轻看黄金规则。

第341页 如我在一个注释中所说明的，诉诸MB5就足够了。如果这些公式有时含有相冲突的蕴含，那我们就不得不在其间做出选择。它们可能发生冲突的时候是：
(1) 我们不能合理地意愿“每个人都以某种方式行动”为真，但(2) 我们能够合理地意愿“每个人都相信，这么做在道德上是可允许的”为真，因为我们知道：如果每个人都持有这样的信念，那也不会有太多人选择依此方式行动。如果在用于如此情况时这两个公式的确相冲突，那么MB5的蕴含会更有道理。为了避免这样的冲突，我们可以从LN5转向

LN6：按照某种方式行动是不当的，除非每个人都能够合理地意愿如下情况为真：在知道不会有太多其他人选择依此方式行动时，每个人都会这么做。

① MM为*The Metaphysics of Morals* (《道德形而上学》)的代码。

但这个公式极类似于MB5，乃至不值得同时讨论这两个公式。而且我认为，MB5不仅更接近康德的观点，而且显然更优。LN6过于简单，因为为什么不会有太多人选择按照某一方式行动，这个问题并非无关紧要。例如，某些人之所以克制自己不按照某一方式行动，是否因为他们相信：鉴于这么做的人数已达相当数量，再这么做将变成不当的。在这一点为真时，依此方式行动的人就可能不公正地受益于其他那些有良知的自我克制者。我们与其如自然法公式所要求的那样纳入关于人们如何做的细节描述，不如纳入道德信念公式所指出的信念内容之细节为好。这个公式本身也更有道理。如我所言，通常在问“如果每个人都那么做会怎样？”是不恰当的时候，问“如果每个人都像你那么想会怎样？”却总是恰当的。

第342页 康德式契约论公式……就把MB5重述为康德式契约论公式而论，我们由关于人们的道德信念的主张转换为人们接受的原则。这样的原则更类似于康德所诉诸的准则。并且这一转换有这样的优点：由于能够将这些原则视同规则或策略，我们就可以通过部分地诉诸它们被普遍接受所产生的效果，而更可行情地评价这些原则。如果以这种方式来评价道德信念，那我们搁置这些信念是否为真的问题就是没有道理的。规则与策略不存在真假问题。但康德式契约论者可以是认知主义者（Cognitivists），因为康德式公式陈述的是某种可能为真的信念。

第346页 ……高契尔拒绝诉诸这样的直觉信念。高契尔还坚持认为，如果我们接受其契约论以及他的最低度道德，那么他就能说明：即便在自利的意义上，不当地行动也不可能是合理的。高契尔宣称，任何其他的道德理论都没有实现这一目标。我在附录B讨论了这种论点。

在运用于道德时……极具吸引力，而且也得到罗尔斯的某些非契约论主张和论证的充分支持。一个例子是罗尔斯关于自然界抓阄的任意专断性（arbitrariness of the natural lottery）的有力主张。我对于罗尔斯的这一评价，是依照内格尔（1973）和巴里（Barry 1989, 1995）的观点。

第353页 ……最大的最小值论证具有未免太过极端的蕴含。即使在运用于社会基本结构时，最大的最小值论证也有这样的蕴含。罗尔斯有时在宽泛意义上界定最不利群体，乃至该群体包括的许多人比其他某些人境遇更好。例如，按照一个提议，最不利者是这样的一些人：其收入低于不熟练工人的平均收入（罗尔斯 [1971] 第98页，RE^①，第84页）。但最大的最小值论证要是合理的，就需要对这一群体做出窄得多的界定。该论证蕴含着：每一个人都应该努力使自己的最差可能收入尽可能得到改善。按照罗尔斯提出的较宽泛界定，我们应当选择的政策，将改善最不利群体中具有代表性的或一般的成员之境遇，即便这更不利于该群体之中的最不利者。这恰恰是在运用于社会整体时罗尔斯所谴责的政策。在辩护他对最不利群体的宽泛界定时，罗尔斯写道：“我们在某一刻有权请求实践的考量，因为对于做出更精细

① RE为*A Theory of Justice* (Revised Edition) (1999)即《正义论》(修订版)的代码。

的区分，哲学或其他论证的潜力必定迟早是会被用尽的”（第84页）。但是，把最不利者群体描述为那些同等最不利的人并不存在任何困难，因为这些人不会比任何其他入过得更好。

第354页 罗尔斯只是拒绝给予……那就绝无可能纳入非效用主义的考虑。在其最后的著作中，罗尔斯表达了对其如下规定（*stipulation*）的怀疑，即我们在无知之幕后“没有任何根据去评估机率”。他写道，“终究说来，要为这一规定辩护，还必须给出更多的说明”（罗尔斯 [2001] 第106页）。但没有做出任何进一步的说明。

罗尔斯补充了一些其他的规定，这些规定容许他给予其关于机率的主张以较小的权重。他要求我们假定：通过选择其正义原则，我们会确保自己拥有“令人满意的”的福祉水准，由此“几乎不在意”企及甚至更高水准的问题。我们还应该假定：如果选择任何其他的原则，我们就会面临境遇差得多的风险。罗尔斯坚持认为，按照这些假定，选择其正义原则对我们来说就是合理的。接着罗尔斯考察这样的反驳：由于补充这些假定，他使得其理论吻合于一个版本的规则效用主义，因为他的原则将是那种其被接受让一般人的境遇尽可能优裕的原则。罗尔斯的答复是：按照他的界定，规则效用主义者并非效用主义者（罗尔斯 [1971] 第181-182页及注释31，RE第158-159页注释32）。这个答复令人失望。罗尔斯早先曾说明，其目标是提供一种理论以作为对所有形式的效用主义的替代。我们如果接受某一观点但给它一个不同的名字，那并没有提供对这一观点的替代。

我尽管已论证罗尔斯的契约论失败，但应该重申，其《正义论》是一部出色之作。

第360页 ……无人能够合情理地拒绝。我们不应该假设：在两个人发生分歧时，其中至少一人必定是不合情理的。有可能存在着合情理的错误。但如果这两人对于对方原则的拒绝都不是不合情理的，那或许就不存在人们不能合情理地拒绝的任何相关原则，斯坎伦公式由此就会失效。因此，当斯坎伦主张无人能够合情理地拒绝某一原则时，他的意思有时应该被认为是：任何拒绝该原则的人由于未能认识到他人的道德主张或予以以充分的权重，而都是在犯道德错误，即便这或许并非不合情理的错误。在此，斯坎伦就会是在某种更为狭窄的意义上使用“不合情理地”。然而，在普通意义上任何人能否合情理地拒绝某一原则，这一点通常是清楚的。

第363页 通过诉诸斯坎伦公式的一个版本，我们可以言之成理地……对此的了解也会使我们许多人感到焦虑。这种焦虑可能不是合理的，但这并不会削弱这样的主张……焦虑与不信任论证。在给出这一论证的过程中，我搁置了斯坎伦观点的一个特征。斯坎伦主张：就对原则的拒绝而论，我们不能诉诸多个人群一起共同承受的利或弊。我们如果遵循这种个人主义限制（*Individualist Restriction*），那就无法通过焦虑与不信任论证来反对行为效用主义关于移植情形的观点，因为该论证诉诸对众人产生的负面效果。我认为，如我在第二十一章所论证的，斯坎伦应当放弃个人主义的限制。

第368页 有一种直截了当且令人完全满意的辩护……在这种意义上使行动为不当的属性或事实。在第八十七节我将对这种区分做进一步的讨论。

第374页 如我在一个注释中所说明的，行动的正当与否，不能取决于……同时提出如下两种说法，是不可能正确的：

某些行动是不当的，因为我们如果这么做就是坏的；

以及

我们如果这么做就是坏的，因为这样的行动是不当的。

不当的行动必定有某种其他使之成为坏或不当的特征。如下也不可能正确：

某些行动是不当的，因为这样的行动为最佳原则所不容；

这样的行动为该原则所不容，因为我们如果这么做就更坏；

以及

我们如果这么做就更坏，因为这样的行动是不当的。

契约论者必须主张：在运用他们的公式时，我们不应该诉诸由行动的不当性所可能提供的道义式理由；正与此一样，后果论者必须主张：在运用其公式时，我们不应该诉诸正当或不当的行动之道义的好或坏（deontic goodness or badness）。后果论者在做出关于“我们行为的正当性如何取决于有关何者最好的事实”的主张时，应该是在我们称作搁置道义价值的（deontic-value-ignoring）意义上使用“最好的”这个词。

类似的主张适用于非后果论者。我们拒绝行动后果论，如果我们相信

（A）某些行动是不当的，即便在人们如果这么做会更好时也是如此。

为了说明（A），我们可能主张

（B）撒谎、偷窃或食言通常是不当的，即便在这些行动会产生更多的善时也是如此。

我们如果相信（B），那也会相信人们如果依此方式不当地行动将更好吗？如果我们相信这样的不当行动其本身就是坏的，那答案可能就是否定的。但（B）不会由此就说明（A），因为我们不会相信：这样的行动是不当的，即使人们如果这么做会更好也是如此。因此，我们如果通过做出类似于（A）的主张来拒绝行动后果论，那就可能需要在搁置道义价值的意义上使用“更好的”这个词。

如某些后果论者所认为的，如果行为可以在道义上是好的或坏的，那我们可能反对说：后果论的理论不应要求我们搁置这些行为的价值。但是，与道义式信念的限制一样，这种道义价值的限制（Deontic Values Restriction）所适用的仅仅是我们道德思考的某些方面。后果论者做出各种如下的相关主张：我们行为的正当性如何，取决于事情如何进展会最好。仅当运用这些主张时，我们才不应诉诸我们关于道义价值的信念。在其他时候，我们可以在不受限制的通常意义上诉诸关于何者为好坏的主张。在本书随后的论述中，我有时会在搁置道义价值的意义上使用这些词；但在大多数情况下这种区分不会产生任何影响。

第375页 按照这种观点……还可能存在许多其他形式的间接后果论。例如，按照动机后果论的一个版本，每一个人持有的最好动机是这样的：该动机被这个人持有将使事情进展最好。我所称的“行动后果论”是运用于行动的直接后果论。但也可

能存在属于间接后果论的行动后果论者，因为他们将后果论标准直接用于行动，但仅间接地用于其他，比如规则或动机。按照这种观点，尽管最好或正当的行为是那种将使事情进展最好的，但最好的规则并非其被接受会使事情进展最好的，而是“总是做会使事情进展最好之事”的规则；而最好的动机不是其被持有会使事情进展最好的，而是总是力图做会使事情进展最好这一动机。（在卡根 [2000] 及 [1998] 第6-7章，这样的各种可能性得到了很好的讨论。）

第393页 在相信“以杀某人作为手段去救其他人”为不当的人之中，大多数……我们就应当将这一事实在某种审慎的意义上视为好事。不时有人主张，我们不可能拥有不偏不倚的理由想要某人行动不当。但情况并非如此。我们的主张应该至多是：我们总是拥有不偏不倚的理由不想要任何人行动不当。我们可以主张：在较小的恶情形中，如果无人杀死任何人以用作手段，那将再好不过；即便那五人会由此死去也是如此。但我们现在假设的情况是，至少有一人会做如此不当之事。你杀我以用作手段，由此做不当之事；这尽管很糟糕，但显然更糟糕的是：格蕾和格林各杀两人以用作手段，由此做不当之事。如果唯一的可能性就是如此，那么其他大多数人就更有理由希望你这么做，因为这会是两恶之中较小者。由此被用作手段而被杀的人会少一些。因此，如果其他大多数人得知，你这么做的阻止了格蕾和格林的不当行动，那么他们应该乐于接受这一消息而认为事情有所改善。这种观点还蕴含着：你会既有想要自己行动不当的不偏不倚理由，也有这么做的不偏不倚理由。但我们可以融贯地认为：我们具有其他的个人相关的（person-relative）道义式理由不这么做，且这种理由能压倒上述的不偏不倚理由。如果这一点正确，那么基于全面考虑，你就有决定性的理由不去做“你有不偏不倚理由去做且我们全都有不偏不倚理由想要你去做”的事情。

第399页 与前提（E）相比，这个前提……因此，如（G）所主张的……在第八十一节我讨论了某些可能的例外。

第400页 一旦与（H）合并……我在一个注释中为这些主张做了进一步的辩护。在《关于德里克·帕菲特〈论重要之事〉的论文集》（苏伊坎南编）（*Essays on Derek Parfit's On What Matters*. ed. Suikannen [2009]）中，对于该论证的较早版本以及我的一些其他主张，好几人提出了反对意见。

在一篇优雅而新颖的论文中，吉迪恩·罗森（Gideon Rosen）对康德式公式提出了多种反驳。在运用该公式时，我们问的是：哪些原则是其被普遍接受是每个人都能够合理地意愿或选择的。罗森论证说，如果我们得知存在某个恶魔或恶意的精灵——若有任何原则被普遍接受，它们就会制造大难——那么该公式就会失败。罗森主张，这一反对不会由于这一虚拟情况的荒诞性而受到削弱，因为道德理论应当适用于所有可能的情况。

上述主张虽然有道理，但不适用于某些类型的道德理论在对道德推理的论说之中所诉诸的思想实验。我们不可能宣称，这样的某种理论应当诉诸一切可能的思想实验。优良的理论诉诸的可能仅仅是某一类这样的思想实验。我的意见是，要答复罗

森的反对，我们可以修改康德式公式。该公式可以诉诸每个人都能够合理地选择的原则，只要每一个人都知道绝不存在这样的恶意精灵：若有任何原则被普遍接受，它们就会制造大难。在考虑这种修改时，罗森反对说，这会降低康德式公式的说服力。我认为，反过来才是正确的。若该公式允许我们设想存在这样的一种恶意精灵，则其说服力会降低。我认为，这同样适用于罗森其他类似的反对。

接着罗森主张：某一道德原则要成为最高的，那就必须不仅告诉我们哪些行动是不当的，还必须描述使得这些行动为不当的最根本的性质。罗森坚持认为，康德式契约论没有实现第二个目标。我同意。按照我的解释，康德式契约论至多能够宣称，其所描述的是某种高阶的致不当特性，它可以纳入或汇聚所有其他的致不当特性。这些其他的特性通常更重要。如我所言，“有些行动所经受的反对，比康德的公式或任何版本的契约论或规则后果论所提供的反对都既更清晰也更强烈”（现可见本书第414页）。

雅可布·罗斯（Jacob Ross）对规则后果论的康德主义论证提出了一些反驳。在陈述这一论证的过程中，我已大为受益于罗斯此前给我提出的若干出色的反驳。在这篇已发表的论文中，罗斯首先指出，某些最优原则可有略微不同的版本，它们为不同的人所接受会好得多。例如，假设有两群人数相等的人，我们可以救其中一群人的生命。一种最优原则可能要求我们救离我们较近的那群人，而另一原则可能要求我们救较远的。按照这些假设，就绝不存在任何这样的最优原则——其被接受是这两群人都能够合理地选择的。这一反驳类似于大赌注的利己主义所提出的反驳，且能够以类似的方式予以答复。

接着罗斯设想有两种原则，其中之一是（1）明显非最优的，因为其被接受会让事情不利得多，但（2）其被接受对于现存的或将要存在的每个人来说都会有利得多。按照这样的假设，每个人就都能够合理地选择这一非最优原则。因此，康德式公式在此就不会要求我们成为规则后果论者。罗斯的例证之一是：两种原则之一允许我们使用廉价能源，这种方式会大大降低较远未来的生活质量。如果我们遵循该原则，那在短期内会通过催生“一种强劲得多的全球经济”而有利于所有现存的人们。我们的行动还可能有益于所有未来的人，因为他们可能既拥有值得过的生活，其生存也受惠于我们的行动。情况有可能是这样的：如果我们遵循另一种原则，亦即最优原则，那么后面生活的就会是不同的未来人，他们的生活质量也会高得多。我认为，罗斯关于这一情况的主张不可能全部为真。我们所有人都有不偏不倚的理由不选择任何如下的原则：其被接受将大大降低未来的生活质量，由此使事情的进展大为不利。而我们不可能有道理去设想：由更强劲的全球经济这一效果所提供的、选择该原则的个人理由，将足以比得上每个人的不偏不倚理由。许多人的福祉并非如此高度地依赖于其收入。我会乐于放弃我的大部分收入，只要由此能够带来这样的结果：人们将按照大大提高后代的生活质量的方式行动。我们不能仅仅为了有利于我拥有更多消费品的缘故而合理地选择某种其他的原则，由此使事情进展大为不利。我认为，类似的主张适用于其他许多人。

我同意如下情况是可设想的：某一原则之被接受，既在整体上使事情进展大为不利，但又大为有利于现存的或将要存在的所有人。例如这种情况可能为真，如果该原则之被接受会给予现存的每个人一千年幸福而有青春活力的人生，但又会终结人类历史，因为这种长寿只有通过某种让每个人绝育的方式才能实现。那么如下情况就可能为真：现存的每个人都能够合理地接受该原则，尽管他们知道事情随后会变得大为不利，因为未来人类将不存在。按照这些假设，康德式公式可能不会要求我们为了使人类能够延存的缘故而放弃如此巨大的个人利益。规则后果论的康德主义论证由此就有必要在某一方面加以限制，因为其结论不适用于此类情况。该论证就至多能够表明：在几乎所有的实例之中，康德式公式要求我们遵循最优原则。

我已讲过，这一限制几乎不会产生什么影响。罗斯写道，如果我能答复由他的虚拟情形提出的反驳，那就会大大加强我的论点即康德式契约论支持规则后果论（第153页）。我认为并非如此。如果我的论证表明，康德式契约论在几乎所有的实例之中都蕴含着规则后果论，那么该论证就是几乎尽其可能地强有力的。我的目标是说明，康德式契约论与规则后果论之间绝无深刻的分歧。如果在几乎所有的实例之中，这两种理论有着相同或类似的蕴含，那么上述论点就是正确的。

在其本卷论文的早期初稿中，迈克尔·奥特苏卡（Michael Otsuka）对于我关于康德主义论证的一些早期主张提出了若干有力的反驳。这些反驳导致我放弃了这些主张，并补充了第五十九-六十一节以及附录C。我既然已放弃奥特苏卡先前所批评的大多数主张，那就没有必要在此讨论这些批评，而只需为它们给予我的极大帮助而对奥特苏卡表示感谢。

在其发表的论文中，奥特苏卡对我保持的主张做了一些评论。根据我的前提（D），最优原则正是每个人都会有最强的不偏不倚理由去选择的。奥特苏卡宣称，（D）要是正确的，我就必须修改我关于何者构成最优原则的解释。就对事情如何进展最好的描述而论，我必须考虑应受尊重的而不是应被促进的那类价值（第62-67页）。但我认为，他提议的这一修改是没有必要的。当存在某种应受尊重的价值——例如，“对人的尊重”的价值即是如此——的时候，尊重这种价值的行动就含有那种应被促进的价值。我们如果这么做就会更好。

奥特苏卡坚持认为，我的论证“有淹没传统上被当作非后果论的道德理论之危险”（第68页）。其中一例是卡姆（Kamm）的这一观点：如果每个人都接受某些义务论的禁令，事情就会进展最好。按照我的解释，如果卡姆的观点正确，那么这些义务论原则就是最优的，且将为规则后果论者所接受。这一事实绝不会对我的论证构成反驳。这些义务论原则并非行动后果论的，因为它们主张：某些行为是不当的，即便我们知道这样的行为会使事情进展最好。但这样的原则可能是规则后果论的。这样的理论要求我们遵循最优原则，即便我们知道由此没能做到会使事情进展最好之事。规则后果论可能接受这样的最优的义务论禁令（另见第六十七节）。奥特苏卡描述了某些其他的方式——由这些方式来看，如果我的论证是合理的，那些被广泛认为是相冲突的观点就实际上不相冲突。如前所示，这绝不构成反驳，因为我所

要力图说明的部分在于：表面上相冲突的观点实际上并不冲突。

对于我关于新理想世界反驳的首次答复，迈克尔·里奇（Michael Ridge）提出了一个略加修正版。我已满怀感激地采纳了这一修改。里奇接着宣称，如果我以某种纳入我的首次答复的方式来陈述康德式契约论公式，那么我的规则后果论的康德主义论证将经受某些反驳。我同意。这是我采取如下处理的一个原因：我没有以此方式来陈述我的论证，并且写道，我将搁置由新理想世界反驳引起的复杂性以待进一步的讨论。

里奇还提问说：如果我们以他提议的方式修改康德式公式，那么我的康德主义论证能否成功？令人失望的是，他没有讨论这个问题，而是代之以这样的问题：罗尔斯式契约论的一个版本要求我们将我们的预期效用最大化，其蕴含是什么。因而我仍不知道，我的康德主义论证能否按照里奇提议的方式修改成功。对于我仍没有尽量答复这个问题，里奇称之为“异常”（第84页）。但我认为，由新理想世界反驳引起的复杂性，不可能削弱我有关康德式契约论与规则后果论两者关系的任何主张。

西尼尔·摩根（Seiriol Morgan）令人印象深刻的论文，对我关于康德的普遍法则公式所提议的修改提出了一些令人困惑的主张。摩根接受我的如下提议：我们应该用对行动者行动的道德相关描述，来代替康德对于行动者准则的指涉。他甚至称这一修改“全然是康德主义的”（第45页）。我还提议，康德的公式不应针对行动者能够合理地意愿什么，而应该针对每个人都能够合理地意愿什么。摩根写道，这一修改没有产生任何差别，因为“实践上合理的意愿是什么……对于每个理性的深思熟虑者（deliberator）来说都是一样的”（第57页）。摩根尽管认为这一修改是不必要的，但无法反对我以这样的、第二种方式对康德公式的重述。没有产生任何差别的修改不可能是不可接受的。我们如果将这些修改合并，就得到了我所称的康德式契约论公式。令我惊讶的是，摩根写道：任何真正的康德主义行动者都不可能赞成这个公式，“因为这么做将是在其意志之中为自己树立一个矛盾”（第57页）。摩根既然没有对我提议的修改提出任何反驳，其评论就蕴含着：任何真正的康德主义者都不可能赞成这种可接受的、对康德公式的修改。这一观点不可能正确。

我认为，摩根真正反对的是我的如下主张：这种康德式公式支持规则后果论。摩根相信我是误入歧途，这不是由于对康德公式的修改，而是由于诉诸某种关于合理性和理由的非康德主义观点。摩根写道，我们不应诉诸关于我们有充足的理由去意愿什么的主张，而应诉诸关于我们能够无矛盾地意愿什么的主张（第47页）。

我的康德主义论证可以呈现为上述形式。在每个人都可能接受的原则之中，有些会是康德式最优的（Kantianly-optimific），其意思是：人们普遍接受这些原则，会使得事情的进展方式是我们能够无矛盾地意愿的。

我的论证可以变成：

- （A）我们应当遵循这样的原则，其被普遍接受是我们能够无矛盾地去意愿的。
- （B）存在某些原则是康德式最优的。
- （C）我们能够无矛盾地意愿每个人都接受这样的原则。

(D) 绝不存在其他的明显地非最优的原则，其被普遍接受是我们能够无矛盾地去意愿的。

因此

我们应当遵循康德式最优的原则。

摩根绝不会反对以(A)来陈述的康德式公式。我相信，他也不可能拒绝承认(B)或(C)。显而易见的是，按照康德主义的假定，存在某些其被普遍接受是我们能够无矛盾地去意愿的原则。这些主张加起来就蕴含着：康德式公式允许我们成为康德式规则后果论者，遵循这样的康德主义最优原则。因此，即使按照摩根关于合理性的假定，这一康德式公式也支持一种形式的规则后果论。我们没有必要问：由于(D)也是正确的，那么该公式是否也要求我们遵循这样的最优原则。

摩根提出，如果康德的公式蕴含着某种形式的后果论，那对于康德主义伦理学来说将是一场灾难。在做出这一主张时，摩根假定，后果论的理论必定给予人们的福祉以最大的权重。摩根写道：“康德所力图明确表达的是这样一种道德哲学：它坚持自由是最重要的价值，当然也就是比幸福更重要的价值。帕菲特的‘康德式后果论’根本没有这么做。他的道德理论反而看来是打算促进另一种价值，一种我们或许可称之为‘福祉不偏不倚主义’(well-being impartialism)的价值。而这种价值一旦如帕菲特打算做的那样被提升到最高价值的地位，那就是一种与康德相抵牾的价值”(第59页)。

我认为，这些评论包含了某些误解。我的论证所诉诸的不是康德式后果论，而是康德式契约论。而我对于规则后果论的康德主义论证，并没有赋予福祉以最高的价值。我们如果相信自由比幸福重要得多，那就会影响到我们关于如下问题的观点：我们能够合理地或无矛盾地意愿事情如何进行。我们不可能意愿每个人都接受的原则是这样的：它导致我们促进福祉的方式(用摩根的话说)“会为了其他的利益而阻碍个人的自由和能动性”(第57页)。摩根的主张，与其说是否认康德式公式支持这样的康德主义最优原则，不如应该说仅仅在于：这些原则会体现一种“康德主义的价值观”，它极其不同于其所称的“福祉不偏不倚主义”。

摩根还写道，如果有人向康德说明其公式支持某种形式的后果论，那康德会认为这是“一种毁灭性的结果，一种会揭示他在道德哲学方面的整个见解都处于混乱状态的结果。因此，帕菲特的论证成功就蕴含着：一位重要的哲学家完全混淆了其自身哲学的性质与蕴含，因为其主要观点实际上所支持的，不是其自身的而是他最为留意去抵制的那种人的伦理见解”(第42页)。

我认为这些主张是错误的。康德写道：“每个人都应当总是努力促进一个由普遍德性与应得幸福组成的世界”。像其同时代的许多人一样，康德认为，通过遵循常识道德的各种原则，我们能够最好地促进这样的幸福。而且如我在第410页所指出的，康德甚至一度提议一种享乐主义版的规则效用主义。摩根的主张与其说是在宣称：如果康德的理论蕴含某种形式的规则后果论，那将毁掉康德；不如说应该是：鉴于康德关于价值的假定，康德本该提出一种既非享乐主义的亦非

效用主义的规则后果论。(对于类似反驳的讨论,见第二卷沃尔夫和赫尔曼的评论以及我的答复。)

詹姆斯·雷曼 (James Lenman) 为其表现主义的 (expressivist)、基于欲望的理由理论做出了有力的辩护,其中他写道:“欲望之所及,即是理由之所及;但仅当欲望是你的。从第三者的视角看,欲望之所及,仅止于欲望自身”(第13页)。我也会这样去描述这种观点。尽管雷曼主张存在着规范的理由,但其观点蕴含着:在客观地看待世界时,我们应该断定绝无这样的理由。雷曼正确地批评说:对于有关我的这一信念——即,存在这样的理由——的形而上学的和认识论的反驳,我没有给予讨论。现在我通过写作第三十一至三十四章,已尽量答复了这些反驳。我在第十一节简短讨论了迈克尔·史密斯颇有力度的论文。在此我要补充说:在其论文的最后部分,史密斯假定,按照我的观点,关于理由的事实可以通过诉诸关于价值的事实而得到解释。情况并非如此。

第406页 这一断言或许看上去无可争辩。而如果该断言正确,康德式公式的这一版本将要求我们成为行动后果论者。在卡根 (2002) 第128页及第147-150页,卡根提出了一个部分类似的论证。许多哲学教授告诉学生,康德的普遍法则公式与行动后果论相冲突。这些人认为,不可能合理地意愿行动后果论的准则成为一种普遍法则。有些学生必定会问:“为什么不能?为什么我们不可能合理地意愿每个人都去做使事情进展最好之事?”我们会指望:如今存在对这个问题的某个标准答案,它可以在伦理学导论的许多课堂上被反复讲授。令人惊讶的是,情况并非如此。就我所知,卡根是首位讨论过这个问题的作者。(而西季威克写道:“我当然能够意愿如下准则作为一种普遍法则:人们应该以旨在促进普遍幸福的方式行动;事实上,它是对我来说完全清晰的唯一法则,是我因而能够从普遍的视角看决定性地意愿的唯一法则” [ME^① xxii]。)

在论证我们能够合理地意愿每个人都遵循行动后果论的准则时,卡根所诉诸的主张是关于工具的和自利的理由。卡根指出,如果选择每个人都成为行动后果论者,我们就可能被要求为了他人的利益而做出重大牺牲。卡根宣称:鉴于这一“逻辑的可能性”即我们可能处在任何一人的境地,承担这种风险就是合理的。这等同于设定某种无知之幕,如罗尔斯的契约论版本一般。在黑尔 (1997) 中,黑尔给出了一个类似的论证。在某些重要的方面,这些论证不同于我一直在讨论的康德主义论证。另一种甚至更为不同的对于后果论的康德主义论证,见卡斯基 (Cumiskey) (1996)。康德的文本具有取之不尽的丰富内容,在不同的人那里激发出极为不同的思想。

卡根坚持认为,我们能够合理地意愿每个人都遵循行动后果论的准则。那么,我可能是简短论证我们不能的第一个作者。

① ME为*The Methods of Ethics* (《伦理学方法》)的代码。

如果每个人总是做使事情进展最好的任何事，那么每个人的行为在大多数情况下会有最好的可能后果。如吉伯德（Gibbard）（1965）和雷根（Regan）（1980）所指出的，这一点并不总是正确。在某些情况下，我们每一个人可能在遵循AC，尽管我们一起并不是在做使事情进展最好之事。对于某一群体的每一个成员来说如下情况可能为真：如果他独自一人采取不同的行动，那就会使事情变得更不利；但如果每个人都采取不同的行动，事情就会变得更有利。一个这样的例子是如第313页^①所描述的失误情形。不仅是在我们都做A，由此救所有人生命的情况下，而且在我们都做B，由此仅救一些人生命的情况下，我们每一个人都已遵循AC。如果我们其中一人做B，那另一人如果做A就会更不利。但如果我们都做B，我们所救到的人就比我们本可能救到的要少。这种复杂性不会削弱我在文中的主张。

第406页 如前所示，就失去诸多深爱、忠诚及个人目标而言，我们许多人将失去太多使生活有价值的东西。这提供了[说明如下观点的]另一种方式：我们不可能总是按照使事情进展最好的方式行动。如果拥有这样的动机——其被持有会使事情进展最好，那我们就经常要选择按照使事情较不利的方式行动。但如果我们放弃这样的动机，乃至从不依此动机行动，那我们就会使得事情在总体上较不利。按照这些假定，通过促使自己成为那种并非总是按照使事情进展最好的方式行动的人，我们将会使事情在总体上进展最好。

然而，该公式的确……而且与UA最优原则相比，这样的原则更类似于AC。这主要是因为，就普遍遵循何种原则会使事情进展最好这一问题而言，我们可以忽略以多种方式出现的如下情况：在努力使事情进展最好时，人们可能由于误算、自我欺骗以及类似的原因而误入歧途。我们还可以指出，按照某些版本的规则后果论，我们会诉诸在我们生活着的这一时段之内为最优的原则。康德式契约论也可以采取这种形式。如果我们要问哪些原则在21世纪是UF最优的，这样的原则会比在其他大多数世纪里甚至更接近于AC。鉴于当今世界财富与权力的极端不平等、贫困的广泛存在，以及我们的技术进步，与以前世纪的其他人所做的相比，许多行动后果论者如今所能做的善事要多得多。因此，AC现在可能是UF最优的。但我相信，AC在以前的世纪里不是UF最优。而且若如我们所希望的，这样的贫困与不平等会得以消除，则AC将不再在未来的世纪里成为最优的。

在此值得提及康德同意原则的一个修正版。按照我们可称之为的普遍可意愿的行动公式（Formula of Universally Willable Acts）：某一行动是不当的，除非该行动能够为每个人所合理地意愿。

每个人都能够意愿某一行动，如果每个人在有选择的情况下能够合理地选择该行动被执行。这个版本的同意原则就其蕴含而言可能也更接近AC。

第409页 要问如何能够最接近康德的理想，我们就必须比较德性之善与幸福。在此

① 此处原文为第000页以上，似为印刷错误。

易误入歧途。有些作者主张：如果不得不在尽义务与促进幸福之间进行选择，那我们就应当总是尽义务。但我们能够接受这一主张，即便我们相信：既然我们唯一的义务就是促进幸福，那就绝无做出这一选择的必要。

第413页 我们还可以宣称（K）……以及（L）……我们可能注意到，这些主张不能倒过来陈述。我们不能做出如下无法得到辩护的主张：如果每个人都能够合理地意愿某一原则是被普遍接受的，这就使得该原则成为其被普遍接受会使事情进展最好的原则之一，由此使之成为最优的。某一原则被接受的效果，并不仅仅取决于人们能否合理地意愿它被接受。我们也不能这样主张：如果某一原则是无人能够合理地拒绝的唯一原则，这就使之成为其被普遍接受是每个人都能够合理地意愿的唯一原则。我对（L）的论证是由其上的（A）至（I）组成的，而且绝不存在这种对于（L）倒转版的类似论证。

第415页 道德理论要取得成功……那我们就可以有拒绝这一理论的正当理由。就我们可以有正当理由拒绝某一理论或信念这一主张而论，我的意思并不是说，该理论或信念是错误的。我们可以有正当理由持有某些错误信念。

第417页 如果上述主张是正确的……就若合符节了。尽管康德式规则后果论有或许相冲突的不同版本，但这样的冲突并不是产生于这种观点的康德主义与规则后果论这两部分之间。

第417页 规则后果论反而可以建立在……根据康德式规则后果论者的观点，我们之所以应当遵循最优原则，是因为它们是“每个人都能够合理地意愿其成为普遍法则”的仅有原则。在此意义上，这个版本的规则后果论建立在康德式契约论的基础之上。然而，如我也讲过的，这些原则正是因为是最优的，才是其成为普遍法则是每个人都能够合理地意愿的原则。在这另外的一种意义上，规则后果论才是更为根本的。但在此毫无矛盾，因为这些观点是以不同的方式相互支持的。还可以指出，尽管康德式契约论为规则后果论提供了更坚实的基础，但只有规则后果论才是能够就其自身被接受的。如我在第六十二节所论证的，只有最优原则才是每个人都能够合理地意愿其为普遍法则的原则。因此，康德式契约论者必定是规则后果论者。

第433页 附录B 合理的不合理性。该附录写于1994年，是为了答复收入高契尔（1997）的文稿。我没有尝试将高契尔后来的著作纳入考察范围。

在一篇早期论文中，高契尔坚持认为……高契尔（1975）。该论证最完整的陈述见MA^①……我力图驳斥这一论证。见一份未刊稿及RP第七-八节。

高契尔认为……我们可以认为他是在诉诸合理利己主义。由于高契尔以我们的效用来意指我们目前（present）在完全知情的情况下深思熟虑的偏好（preferences），其所诉诸的严格地说是慎思理论（Deliberative Theory）。但如高契尔所讲的

① MA为高契尔著作*Morals by Agreement*（《协定道德》）的代码。

(MA第6页),他的大多数主张同等地适用于合理利己主义。而高契尔经常使用像“收益”(benefit)和“益处”(advantage)之类的词语,这些词所指的,与其说是我们目前的偏好,不如说是我们的利益而显得更自然。因此,我们在此可以忽略慎思理论与合理利己主义之间的区别——尽管它们的区别通常是很大的。可以姑且假设,在我们讨论的所有情况下,我们目前深思熟虑的偏好与符合我们利益的东西是吻合的。

如果我们诉诸合理利己主义……或对于我们来说是可预期地最佳的。预期最佳的东西,未必是我们能够预期是最佳的东西。有些行动对于我们来说是可预期地最佳的,尽管我们可以肯定地知道,它们实际上不会对于我们来说最佳的。鉴于风险的存在,努力做实际上最佳之事或许是不合理的。

第434页 如我所指出的……那对我们来说甚至更有利。见RP第七-八节。

高契尔的答复是……在MA(尤其是第173-174页)中,高契尔给出了这个答复。在高契尔(1997)中,他后来放弃了我们不可能骗得了他人这一主张。他提出,如果我们仍旧是自利的而仅仅是显得可信赖,那对我们来说会更不利。因而他写道:“可以合理地预期:能做出真诚的许诺所得的总收益……会超过人们在许假诺本可以幸免于罚的情况下许诺所付出的总成本”(第26页)。但是,如果我们不诚实能够侥幸逃脱,那么能够真诚许诺的收益何在?高契尔可能像休谟一样,诉诸心灵安宁和良知健全的收益。但这对于其目的来说似乎是不充足的。高契尔还声称,我们即使一般来说是可信赖的,也能够做出某些不诚实的许诺。但这仅仅是限制了诚实的成本,并没有提出有何收益。我认为,高契尔的独特论证要顺利展开,就需要其早期的假设,即我们不能合理地希望去骗人。

第435页 在对其观点的最初陈述中……例见MA第六章……我对(A)的质疑如下……见RP第七-八节……但是,依此习性行动对于我们来说可能不是合理的。我还推测,改变我们关于合理性的信念可能是合理的。这也旨在有助于高契尔的论证。如果不改变我们的信念,我们就会是在做我们相信是不合理之事,而这可能足以让我们的行动成为不合理的。但在此我们可以忽略这一点。

高契尔一度接受……如他写道(像维多利亚女王一样),“我们是无动于衷的”(MA第185页)。

第436页 根据(B)……高契尔做出断言(B)——他称之为其“二阶承诺”——于高契尔(1997)第40页。我讨论了一个类似的主张,我称之为“(G1)”(RP第13页)。根据高契尔的二阶承诺,按照某一习性行动是合理的,“只要人们合情理地预期,过去和未来坚持该习性都是最有益的”。看来这个主张可能意指,“如果某人既合情理地相信,‘过去坚持该习性是有益的’,又合情理地预期将来坚持它也会是有益的”。但这不可能是高契尔的意图所在,因为它将消除其二阶承诺与一阶承诺(对此的讨论见后)之间的区别。高契尔的意思必定是:“如果某人能够合情理地相信:将过去和未来合起来考虑,获得该习性在其作为整体的人生之中是有益的”。高契尔从(A)到(B)或者说从其三阶承诺向二阶承诺的转移,几乎无损于他对

合理道德的辩护。按照在MA中得到辩护的观点，道德的限制要对我们具有合理的影响力，接受这些限制就必须本该对我们是可预期地最佳的。按照高契尔经修正的观点，这些限制要有合理的影响力，就必须也为人所知地不会总体上不利于我们。高契尔契约论道德的大多数限制都满足这第二种要求。

我们可以修改这个例子……早在此前的某个时候就已死去。或许我本来会遵守某个被证明是致命的命令……根据(B)，我应该拒绝给你椰子而被炸死。可能有人反对说，我获得的习性过于粗糙。或许我应该变成倾向于无视威胁，除非在这样的情况下：我相信依此方式行动将是灾难性的。但是，如高契尔所言，“我可以合情理地认为，[对我的习性]的任何限制都会降低其事先的(ex ante)价值，由此无限制的无视威胁会赋予我最佳的生活前景”（高契尔[1997]第39页）。我们可以补充这一假定：只有无限制的习性才会实际上最有益于我。（不允许该习性呈现为有限制的形式，尚有另一种理由。我们如果这么做，就不得不允许对可信赖的习性加以类似的限制。如我们所见，这会削弱高契尔的论证。）

在运用于承诺时……而继续做一名无视威胁者？高契尔赞成这样的潜在威慑者的行为：一旦威慑失败，她就会灾难性地执行其威胁。他写道，“她坚持到底的理由……只不过是：其失败策略的预期效用取决于她坚持到底的意愿”（高契尔[1984]第489页）。那又如何呢？她的预期可能以前一直取决于这样的意愿。但是，为什么现在她还应依旧忠诚呢？（我们在此讨论的并非真正的忠诚[fidelity]。）

为了应对情形的这一版本……对我来说合理的是设法放弃这一习性。请注意：在这样主张的过程中，我们不必诉诸合理利己主义。我们不必假定：这种企图之所以是合理的，是因为它有益于我。由于高契尔拒绝合理利己主义，那样做就是在回避问题。但即使按照高契尔的理论，设法放弃这一习性对我来说也是合理的。假设一旦我的习性变成灾难性的，我就放弃之。拥有这种元习性(meta-disposition)将符合我的利益。因此，根据高契尔的理论，按照这种元习性来行动现在对我来说就是合理的。

为了应对情形的这一版本……故意招致杀身之祸对我来说就是不合理的。先假设：我如果努力的话，就能够不再成为无视威胁者。如我刚刚论证的，此时保持我的习性对我来说是不合理的。高契尔如果接受这一结论，那还能够主张(B)吗？即便保持我的习性现在对我来说是不合理的，依此习性行动必定仍是合理的——他能够这样宣称吗？

可能存在某些情况，其中尽管保持某一习性是不合理的，但依之行动仍是合理的。例如，假设保持精明对我来说是不合理的。如果我的确不合理地保持这种精明，但“依之行动而去做最有利于我的任何事”仍可能是合理的。然而(B)是一种强得多的主张。根据(B)，即便保持某一习性现在是不合理的，但依之行动必定仍是合理的，其原因仅仅在于：它曾经给予我大于目前成本的收益。我认为，这一主张不可能正确。如果保持该习性是不合理的，那么为什么如果我实际上保持并依之行动

就必定是合理的呢？

如果不合理地保持精明，对于为什么依之行动可以是合理的，存在一种不同的解释：这么做对我更有利。这种行动的合理性要得到辩护，并非必须通过诉诸该习性的或我保持依之行动的习性的合理性。这种情况相当不同于我以自知是灾难性的方式无视你的威胁。如果这样的行动被宣称为合理的，那只能是诉诸我在依之行动的习性的合理性。而如果保持该习性现在对我来说是不合理的，那似乎就毫无理由去断定：依之行动对我来说必定是合理的。

接着假设我尽管做了努力，但不能放弃我的习性。高契尔可能会说，如果这一点为真，那保持该习性对我来说就不是不合理的。它不是我做得到的事。但是，如果我能够放弃，那保持它对我来说就会是不合理的。这似乎就足以削弱“依之行动必定仍是合理的”这一主张。

第437页 高契尔的观点如果诉诸(C)……(C)是对高契尔所谓其观点的“最弱”版本的或其所谓一阶承诺的一个解释。他写道，按照这种观点，人们应该按照某一习性行动，尽管其行为是“代价颇高的……只要他们合情理地预期，坚持该习性在将来是最大限度地有益的”（高契尔[1997]第39页）。

在谈及对该习性的“坚持”是有益的时，他的意思必定是持续拥有该习性。如他所同意的，依之行动可能是代价颇高的。我还要将“坚持”的意思解作“目前坚持”。尽管高契尔可能意指“现在和未来坚持”，但这会使其主张变得不那么有说服力。它不会涵盖这样的情况：起先获得继而放弃某一习性，将是有利的。（假设尽管获得某一持久的习性比根本不获得确实要更有利，但是，仅在一段时期拥有它才是可预期地最佳的。获得这种持久的习性，由此就不是如高契尔所要求的那样“最有益的”。）

第438页 我认为，他并没有……即便这是暂时性的，仅足以违背承诺。高契尔可能拓展他关于可识破性(translucency)的主张。他可能会说，我们不可能有理由去相信：我们如果违背承诺，那么还能够隐瞒该事实。但这样的答复将失去高契尔观点之中的新颖之处，因为这样会归于“诚实总是最佳策略”这一古代主张。

高契尔可能试图给出另一种答复。他可能主张……没有能力去放弃或克服这一习性。按照某一种解读，这个主张必定正确。或许有人说，我们如果能够中止我们的习性，那就不是真正可信赖的。但这种解读是不相干的，因为对于高契尔的目的来说，要紧的仅仅在于我们是否表现得(appeared)可信赖。如下的主张是相当无说服力的：我们如果违约，那就无法在此前——即便我们当时的真实意图是守约——表现得可信赖。

上述主张如果要有助于高契尔的论据，他就必须在其观点之中做出其他的修改。他写道：“如果在人力能及(humanly possible)的范围内，拥有某一习性将导向一种与拥有其他任何习性同样如意的人生，那么该习性就是合理的”（高契尔[1997]第31页）。诉诸人类的可能性(human possibility)似乎与高契尔观点的其他部分相抵触。在别处他宣称，我们不应该问哪些习性是一般地合理的，因为答案可能取

决于特定个人的处境。因而他写道，“绝无必要存在这样一种习性：它独立于行动者的处境而足以保证他的生活尽可能如意；而由此我就没有必要去设想：需要有某种单一的、最为合理的习性”（高契尔 [1997] 第31-32页）。一个人的处境当然能够包括对于这个人来说的一切可能情况。

诉诸人类的可能性也对高契尔的论证提出了一个问题。可信赖不是人力能及的那些东西之中最有利的习性。更为有利的是，表现得可信赖但实际上是精明的；而且这对于某些人来说当然是可能的。高契尔如果诉诸人力能及之事物，那就不得不将可信赖判定为一种不合理的习性，甚至在它为这样的人——由于他们骗不了他人，它是在可能范围内最有利的习性——所持有时也是如此。

第439页 我认为答案是否定的。依据我前面给出的理由……见第260-263页……但无法表明我的行动是合理的。在“应当”蕴含着“能够”这一信条之中，“能够”的含义与决定论（determinism）是相容的。如果否认这一点而我们又假设了决定论，那我们就不得不主张：每一种行动都是合理的。

重新考察先令情形……由此对我大为有利。我如果仅仅是表现得精神失常，那当然甚至会更好。但我们可以设想这是不可能的，因为如果我没有服药，这个抢劫犯就会知道这一点。（或许该药物的一个效应是某种独特的眼神。）确实处于这种状态由此就是对我最佳的习性。

第440页 高契尔反对说，我的疯狂行动……见高契尔（1997）第37页。

我的行动尽管……按照高契尔的观点，这就足以使这些行动成为合理的。当然，其条件是这些负面效果不会超过我的习性之好效果。高契尔没必要声称，如果我自杀或杀我的孩子，那也是合理的。

休谟有个著名的主张……按照高契尔的观点，它们仍然可能是合理的。或许有人认为，从某一方面说，高契尔的观点没休谟那么极端。尽管我的行动有负面效果，但它们必然会被我拥有该习性而产生的好效果所压倒。但在此我们可以回顾：按照高契尔的观点，我如果满足我目前深思熟虑的偏好（这些偏好不必与我的利益相一致），那就实现了我的效用最大化。如同休谟的观点，这些偏好极尽疯狂之能事。两种观点之间的区别在于：按照休谟的观点，我的行为要是合理的，那我就必须至少努力去实现我的目标；而按照高契尔的观点，我的行为只需是某一状态——拥有该状态会实现这些目标——的连带效应。

高契尔认为，我们要是理性的……在此可将该观点称为G。如高契尔所言：“我们的论证将实践的合理性等同于在选择习性的层次上的效用最大化，并且在评价特定选择的合理性的过程中贯彻这种等同性的蕴含”（MA第187页）。

第441页 在我们所关注的情形之中……会对我们不利。看来可能是：如果这一点正确，那么违背承诺就不可能对我们更有利。但情况并非如此。这种负面效果的来源，并非我们对这些承诺的违背，而是这一事实：我们既是可识破的，又倾向于一旦违背承诺对我们有利就这么做。

我们无法在两个层次都追求最大化……将超过我们行为的不利效果。值得解释其原

因。在评价我们习性的效果之好坏的过程中，我们会纳入这些习性将要或可能导致的行为。如果拥有某一习性对我们来说是最佳的，那么尽管这会导致对我们不利的行为，但这样的效果必然会被抵消。对我们习性的评价既然包括对我们行为的评价但不止于此，那就是一种告诉我们综合起来考虑什么对我们最有利的的评价。

高契尔主张，鉴于这一事实……MA第170页……该假设可能不正确……E认同于G的观点：我们应该努力拥有这些习性。可能有人质疑：G是否会建议我们尽可能养成这样的习性。这并不能从如下事实推出来：我们如果这么做就对我们更有利。如果G没有建议我们这么做，那就是对G的一种反驳，并且又会破坏高契尔的论证。但高契尔可以主张：在努力养成这些习性的过程中，我们将是在按照一种有利的或最大化的元习性行动。

高契尔可能认为，我们如果接受E，就总是会做E主张为合理的事。他会承认，实际上我们极少人始终是理性的。但他可能主张，就评估这些理论的可行性而论，我们应该考虑，如果总是做这些理论建议我们所做之事会怎样。由此他可能主张，如果完全遵循S，我们就会总是在行为的层次实现最大化。

高契尔可能认为……我们如果转而养成这样的习性，那就无法总是以其他行为来追求最大化。可能有人反对说：如果我们不可能总是做E主张为合理的事，那么E就无法主张我们应当这么做。“应当”蕴含着“能够”。但这将两个问题混淆了。当我说我们无法总是做E主张为合理的事时，我的意思是：这不是因果地可能的。在比较我们拥有不同习性的效果时，这种可能性是不相干的。“应当”所蕴含的那种“能够”，其含义并不像高契尔所认同的那样要求这种因果的可能性。因为“能够”的这种另外的含义与决定论是相容的。

既然无法总是做……所做之事是否合理。看来可能的是：如果我们不可能总是做E建议我们做的事，那就无法预测我们何时会遵循S。情况并非如此。假设我们现在总是倾向于做我们认为是合理的事。如果知道我们能够养成最大化的习性，那么我们会这么做，即便我们知道这会使得我们后来做出不合理的行为。根据E，养成这样的习性就是待做的合理之事。仅当养成这些习性之后，我们才会开始以E主张为不合理的方式行动。

第442页 在讨论核威慑的时候……见高契尔（1984）和（1985）……我们就不能产生该意图。见高契尔（1985）第159-161页。这样的主张是没有道理的……我们或许由此能够径直决定就这么做。见麦克伦南（McLennen）（1988）。

这可能只是转换了问题……由此使自己获得这种信念就可能符合我们的利益。就可信赖以及高契尔主要关注的习性而论，这样的主张是相当有道理的。如他所认为的，如果无法隐瞒我们的意图，那我们如果打算守约，就可能对我们更有利，即便是在这样做对我们更不利时也是如此。除非我们有这种意图，否则他人就会将我们排除于有利的协议之外。而我们要能够形成这种意图，那可能就不得不相信信守这样的承诺是合理的。

第443页 卡根提出……在给我的一封信中。

高契尔后来修正了他的论证……见MA（第182页）以及高契尔（1997）第31页。而另见MA第170页和158页。

第444页 然而，高契尔正当地拒绝了（D）……见高契尔（1997）第36页。

高契尔能够收回这一主张转而诉诸（D）吗？在某种意义上，高契尔近乎接受（D）。他引证本书中的版本（D）——在他那里被称作“（G2）”——并写道，“在此意义上，我接受……（G2）”。高契尔（1997）第40页。

这一论证可能揭示了某些东西……这个信念是某种有用的幻象。高契尔可能答道：规范信念不是真正的信念，真正的信念可能是真的，或者是幻象。但这并不会挽救高契尔的论证。即使按照非认知主义的（*noncognitivist*）观点，我们也必须给予规范信念这一概念以某些内容。我们必须能够宣称某一行动是合理的，并且能够坚持或拒绝不同的理论。我的意见可以用这些方式来重述。

第445页 刘易斯拒绝……见刘易斯（1985）……高契尔在其答复中否认……见高契尔（1985）第159-161页。

第446页 高契尔提出了支持其观点的又一论证……见高契尔（1997）第30页……高契尔宣称，任何理论都不可能避免这种不受欢迎的后果……高契尔（1997）第36页。

高契尔还提出了另一种主张……‘合理性’……是某种技术性的用语……高契尔（1997）第38页。

第447页 高契尔宣称，任何其他的道德理论都没有实现这一目标。MA第17页。

第450页 这些人相信，这种行动会是不当的。……借助杀死其他某人来救这些人会是不当的。这可能不是对于“什么使得这些行动为可允许的或不当的”最佳描述。对此的另一种解释，见卡姆（2007）第一至五章。

这些注释谈及参考文献。有些注释仅给出了文本段落的开头语，因为这就足以使得参考清晰明了。

xxxiii 康德是自古希腊以来……最好的伦理学著作……我在这些意见上追随 Broad (1959) 143-4。有关西季威克的最好的书是Schneewind (1997) 和Schultz (2004)。

xxxiv 批判哲学必定……*Declaration concerning Fichte's Declaration concerning Fichte's Wissenschaftslehre, Kant, Correspondence*由Arnulf Zweig翻译和编辑 (Cambridge University Press 1999) 560。

这本书没解决任何问题……Sidgwick, HSM 284。

总是考虑周到，常常是精致的……HSM: 177。

西季威克还谈到他“冗长、难懂且乏味……(2006) 398。

西季威克这里指的是另一本书，但我认为，他也会把此主张应用于《方法》一书。

对于西季威克的枯燥……没看过任何伦理学的书。Sidgwick (2000) xxviii。

xxxv 在剑桥……在大路上游动。HSM: 396。

无疑……是雪上加霜。HSM: 92。

我养尊处优……我会是什么？HSM: 170-1。

xxxvi 假定……就是荒谬的。ME 298-9 (我加的强调)。

……由此义务的秩序……ME 第一版 (1874) 473。

我不能基于这种资源……哲学的狂热之途。ME 507 note。

xxxvii ……自私者会错过……微不足道的片段 (ME 501)。西季威克一如既往地补充道：“然而，我并不认为我们能有证成地把前一段落所承认的观点陈述为普遍地真的。有些彻底自私者至少比大部分无私者看起来要幸福；还有些例外的性质，其

- 主要幸福看来派生于活动，实际上是冷漠的，但指向幸福之外的目的。”
……某人说“邪恶啊……坦率而一贯的邪恶。Sidgwick (2000) 118。
西季威克提醒……有点枯燥和让人讨厌。ME 295。
一切强烈的亲密情感……那么令人遗憾。ME 437。
- xxxviii 人们可能说孩子……人们的感恩之心。ME 248注释。
A比B更幸福……任何不义之处。ME 284。
因此谨慎地表达……这样才是有利的。ME 490。
真正具有穿透力的……无法保持的。Mind, 1877 125-6, 引用在Schultz (2004) 349。
[这本书]……假模假样的公平。HSM: 74。
- xxxix 一直在读……归功于它。HSM: 421。
西季威克……对其对手如此公平。
我将尽可能是赞扬它……对伦理学极为沮丧。(1906) 411。
不停地提炼……所获无几。Broad (1959) 144。
- xl 西季威克在再次批评……越来越低。(1906) 93。
这些话也过于严厉……近乎天才的良好感觉。Rashdall (1892) ……极精确的讨论，Broad (1959) 14。
- xli 还存在着更深刻的问题……最令人气愤的”一本。O'Neill (1989) 126。
- xlii 如果我早知道……很不可靠的。Kemp Smith (1915) 531……尽其所能做最好的事情。Kant: R 72。
- xlii 如果我早知道……伟大思想家中最不精确的。Kemp Smith (1915) 527。尽管这话有关康德的《第一批判》，但我认为，它也适用于康德有关伦理学的书。
康德写过……最大的义务。C2 24。
我们如果回顾……自身意愿的方式行动。G 432。
- xliii 自杀……斯多葛学派有关自杀……LE 127,148, 369。
正是“假定”……一个评价者表示……Korsgaard (1996) 126。
- xliv 对于其他的此类例子……纯粹的物件还要低。MM 429-30。
康德有时候……理性批评的宣判。HSM: 177。
我认为，康德的有些观点……以及多样的言说方式。Rawls (2000) 18。
- xlv 西季威克早已离世……对我的鼓舞最大。参见Nagel非常精彩的*The View from Nowhere* (Oxford University Press, 1986)，特别第8章和他的*The Last Word* (Oxford University Press, 1997)。
- 第31页 我们很难解释……算作支持……我在这点上追随Scanlon, *What We Owe to Each Other*, (Scanlon 1998) 第1章。
- 第36页 许多人主张，……这并不是最好的看法……见第16节。
我们若想是完全理性的……这些要求引出了几个有趣的问题。参见Kolodny (2005)，Broome (即出)……Scanlon (即出)……我将在后面给出……In Section 17。

第38页 男管家整个上半身就像冰冻了一样。Wodehouse (1952) 93。

第39页 斯坎伦主张……不可能给予我们理由。WWO 97。斯坎伦称此为推卸责任的观点。

我认为这种观点……是误导性的。尽管斯坎伦主张善性与恶性并不是给予理由的属性，他有时候提及这种派生理由。例如，斯坎伦写道：有不止一种理由要对痛苦中的人做出回应：他的痛苦是坏的，我们有义务帮他减轻它”（WWO, 181）。斯坎伦会同意，第一种理由的根源是使得这个人的痛苦是坏的那些特征。

第46页 大多数事情是好的……就在那，闪闪发光。Korsgaard (1996) 225, 278。

第48页 尽管我们若有理由……通常能给出理由。我在这点上追随 Scanlon: WWO 18–22。

第52页 目的式理由……成为外在地好的。我得出这些区分在部分程度上遵循了 Korsgaard (1996) Chapter 8。

第54页 这些说法是有争议的……无论是它们还是我们的意识状态，都不是坏的。对于一种不同的观点，参见Rachels (2000)。

第55页 首先，许多人相信……Korsgaard (1996) 262。

科丝嘉的话……而赋予其价值。Korsgaard (1996) 284。

第57页 没有人对时间具有这种态度……我讨论了这些时间偏好，见RP的Sections 62–70。（在那个折磨人的讨论中，我并没有交待清楚，在我看来，最理性的态度是时间中立性。）

第61页 知情欲望理论……才是我们应该满足的。我在这点上遵循Peter Railton, ‘Moral Realism’, in Darwall (1992B) 142 and note 15。

第65页 理由的主观主义……Korsgaard (1996) 317。Williams (1985) 19。

我认为，我们应当……如前一样，参见Scanlon WWO Chapter 1。也参见Raz (2000) Chapter 2。

第二……欲望与信念之间做出区分。正如斯坎伦所写：“如果具有一种欲望涉及到明白是欲望之外的事物提供了一种理由，这可以解释‘欲望提供理由’这种观念的可行性。情况的确是‘具有一种欲望涉及到认为具有一种理由’。这种错误在于混淆理由与个人把它看作是一种理由。”（Scanlon (2002) 338）。

第66页 第六……给予完全同等的权重。我在这点上追随Scanlon (2003B)。类似的主张也适用于成本收益分析。这些分析能正确地诉诸人们的偏好，而不用由此假定一种基于欲望的理由理论……例如，诺齐克就主张……Nozick (1993) 176。

第70页 没有那门科学……LE, 58–9 (27: 264–5) 和Sidgwick的 ME 74–5。

第73页 在做出这些主张时……子女或其他近亲的爱。对于这种给出理由事实的讨论，参见Kolodny (2003), (2010) and (即出)。

第77页 主观主义者可能会回应道……“适度的审慎”。Williams (2006) 111。

第78页 然而，主观主义者不能……我们都是程序上理性的。Williams做出了这个区分，见‘Internal Reasons and the Obscurity of Blame’载于 Williams (1995)

36-7。

知道人们是理性的……Rawls (1996) 49。

第79页 也许有人会反驳道……Michael Smith ‘Desires, Values, Reasons, and the Dualism of Practical Reason’, in Suikannen (2009)。我在此总结了斯密所称的R原则(120)。

史密斯写道……但周二除外。(op.cit. note 5)。

第80页 为了阐明这个要求……Smith (2004) 269-70 (原书中的88页是错误的)。

第94页 他们也不能融贯地断言……更有理由要去努力满足的。对于此类理论的类似反驳, 参见Enoch (2005)。

有必要确定特别合理的目的。Korsgaard (1996) 261。如果我们没有评估我们的选择对象, 那么我们的选择是否应被称作合理的, 这就是不清楚的。

第96页 我认为, ……什么对我们是重要的就够了。Frankfurt (1988) 81, and 91 note 3。

第97页 某人真的完全不关注……Frankfurt (2004) 22。

第98页 不是有关关注的事实性问题……Frankfurt (2004) 28。

爱对于爱人者本身……Frankfurt (2004) 37。

第99页 当某个目的……Frankfurt (2004) 56。法兰克福使用“义务”这个词, 但他不是在讨论道德, 因此, 这个义务最多也只是一种决定性理由。

第103页 个人的善……Rawls (TJ) 395。罗尔斯的善的厚重 (thick) 理论令人吃惊地是类似的。

会采用的……Rawls (TJ) 417。

尽管这样的问题……但个人在这种慎思之后事实上会选择什么却是一个心理学问题。正如西季威克提到的, ME 112。罗尔斯声称在给出这个定义时, 他是在追随西季威克。虽然西季威克提出了类似的定义且主张它具有某些优点, 但西季威克拒绝了这个定义, 在部分程度上是因为它不是规范性的。西季威克然后定义其利益为“如果我的欲望与理性和谐一致的话, 假定我只考虑自己的存在, 我实际上应该欲望什么”(ME 109-113)。(在早前的一个版本中, 西季威克把“由理性所规定的终极目的作为我们应该寻求或指向的事物”。[ME 5th edition 112] 我加的强调。)

它是个人在具有……Rawls (TJ) 408 (我加的强调)。

无论我们具有……想要什么或赞同什么。Rawls (TJ) 184-5。

第104页 人们可能会主张……没有任何办法可以超越慎思理性。Rawls (TJ) 560。

第104页 然而, 罗尔斯的主张……我们想要说的一切。Rawls (TJ) 401, 也可参见 111 and 451。

第107页 这些黯淡的观点……大部分主观主义者认为这是理所当然的……甚至对于声称质疑基于欲望的主观理论的某些作者来说, 情况也是如此。例如, 诺齐克提出了超越工具的且是基于人们认知合理性论说的二十三条建议。(Nozick [1993] Chapter V. 但没有一条建议含有这样的观点: 我们具有欲望的理由中, 有些是这些

对象或我们想要之物的内在特征所给出的。

第108页 有些主观主义者的论证……这个理由就是基于欲望的。例如，这个论证是由 Williams 的评论所提出的，载于 ‘Internal and External Reasons’ (Williams [1981]) 102 and 106–7, and in ‘Internal Reasons and the obscurity of Blame’ (Williams [1995]) 39。对于此论证更长的讨论，参见我的 ‘Reasons and Motivation’ Parfit (1997)。

第110页 哲学自然主义者……Darwall (1992) 168。

基于最深刻的反思……Stephen Darwall, Allan Gibbard, and Peter Railton, in Darwall (1992B) and Darwall (1996) 176–7。

我认为，理由的这种自然主义论说……必须是错的。我在这些话中遵循 Nagel (1986) and (1997)。

第113页 许多人……休谟表示……休谟写道，尽管欲望不可能严格意义上与“人的理性相悖”，但它们在“立基于错误的假定”之时可能在不严格的意义上是“不合情理的”。Hume’s Treatise, Book II, Part III, Section III。我在第111节对休谟的观点做了进一步的探讨。

第113页 我们若想是完全理性的……探讨这些要求。参见 Kolodny (2005), Scanlon (forthcoming) and Broome (forthcoming)。

第118页 努力获得真理……实践理由和认知理由支持对不同问题的答案。 See Kelly (2003)。

第119页 然而，正如以前一样……正如斯坎伦提到的…… WWO Chapter 1。

第121页 当斯坎伦讨论……WWO 29–31。

第123页 我已经拒绝了……斯坎伦在某种意义上…… WWO 25–30。

第125页 基于一种普通的观点……这将是一种合理的不合理情况。对于一些例子，可参见附录A和B。(原文为124页是错的。)

第126页 按照其他某种类似的观点……布兰德特表示，参见 Brandt (1979) and (1992)。

第128页 当这种偏好依赖于……在理论上是有趣的……对于一些极迷人的论证，参见 Temkin (1987), (1996), (2009), and (即出), and Rachels (1998)。这样的论证一旦成功就极具实践重要性。

第130页 理由的各种客观理论……我将论证，情况不是这个样子的。参见 Chapter 22。

第131页 在其乏味的大作《伦理学方法》……西季威克尽管称利己主义为“伦理学方法”之一，却是在讨论他称之为个体行动的合理目的的观点。(ME xxviii, 我加的强调。)

实践理由的二元论……ME, Concluding Chapter。这只是西季威克的部分观点。他还做出了其他的主张，我在20节会转向它们。

第133页 西季威克对其观点的捍卫……自身福祉的强理由。用西季威克的话来说，“否认人与人的区分是真实且根本性的，这有悖常识，‘我’由此关注自己作为个体

的生活质量，在某种意义上是具有根本重要性的，且无法在这种意义上去关注其他个体的生存质量：由于这样，我无法明白，我们如何能够证成这种区分不被看作是根本性的，从而决定一个人理性行动的终极目的”（ME 498）……伦理学的根本事实。Findlay（1961）p 294。比较罗尔斯的主张：“效用主义者并没有严肃地对待人与人的分立性”（Rawls [TJ] 27）。

西季威克的二元论……内格尔所称的……见Nagel（1986）特别第8章和9章和Nagel（1991）第2章。西季威克论及“关于人类个体的不可避免的双重观念：作为一个整体的本人，以及作为更大世界的一个部分。有些事物是他把自己作为一个独立单元考虑时是理性地欲求的，还有事物是当他采取一个更大整体的视角时，他必须再次承认要理性地欲求的。”（第三版的ME, p 402, 引自Schneewind（1977）369。）西季威克还写道：“从整个宇宙的视野看，任何个人的利益不比任何他人的利益更为重要，作为一个理性存者我的目标必然在于普遍的利益，而不仅仅在于它的一个特定部分。”（ME 382）（内格尔称此为“对个人自己视野的超越……伦理学中最重要创造性力量”[Nagel（1986），8]。）

第133页 我们接下来假定...没法给我们任何指南。ME 508。

第134页 我们能称此为两视角论证。西季威克并没有明确地断言（D）和（E），但他的推理看来需要这些前提。

第136页 我很快会转向……关注他们的福利。对于这些理由的讨论，参见Kolodny（2003）。

假定我被从……Nagel（1986）160。

第138页 在思想中，痛苦可以与……Nagel（1986）161。

第142页 西季威克怀疑……ME 386 note 4。

有的作者考虑……就是傻瓜。Reid（1983）598。里德可能没有承诺这种观点，因为他认为我们并没有面临这种困境。

第143页 我们对于行动之内在合情理性……ME第一版（1874）473。既然西季威克在后面版本中删除了这一段，这里值得整段引用：“但是，理性利己主义原则和建构整体的义务体系之原则之间的这种根本对立，在与其他方法调和之后只会显得更为清晰与尖锐。‘履行社会义务有益的不是自己而是他人，这个古老的、不道德的悖论无法为经验论证所完全驳倒：我们越是研究这些论证，就越是被迫承认：如果我能依赖的只有这些，那么就必定存在着这种悖论为真的情形。然后我们就只能与巴特勒一道承认：寻求个人自己的幸福终究说来是合情理的。因此，如果没有一个可为经验证实且调和个人和普遍理性的假设，没有（不管以何种形式出现的）这一信念：我们以为在现实世界里没有完善地实现的道德秩序实际上依然是完善的，那么我们对于行动的内在合情理性的整体信念就会坍塌。如果我们拒绝这种信念，我们也许仍然能在非道德的宇宙中发现思辨理性的恰当对象，能够在某种意义上最终得到理解。但是义务的秩序由此就会被归为混乱：人类知识分子建构一种理性行动的完美理想的长期努力看来就注定是要失败的。”

- 第144页 还有第三种可能性……去行正义之举。Rawls (TJ) 575。
- 第146页 我们接下来要提到……没有那么重要。例如，这个得到了强有力的论证，参见Kolodny (2005), Scanlon (2007), and Broome (forthcoming)。
- 第156页 我们很难明白……更讲不通。参见 ‘Moral Luck’ in Nagel (1979)。
- 第158页 在所有使得……在逃的奴隶逃脱。参见Bennett (1974) 123-134。
- 第161页 我们还有一个理由解释……正如西季威克同样指出的……ME 207-8。
- 第168页 从宇宙的视角看，任何单个人的利益……ME 382-3。
- 第168页 我们最好是把这类后果论……“卑鄙的”的目标。ME 200, 403。
- 第169页 这些人可能认为……道德的外部对手。例如，西季威克写道：“效用主义道德可能为选择‘普遍利益’作为其终极目的那些人接受，无论这些人这么选择是基于宗教根据，是由于其心中不偏不倚的同情之主导，是因为他们出于良知的行动吻合于效用主义原则，还是出于如此种种的理由之某一组合。”（Sidgwick [2000] 607）西季威克对第二和第三种群体的区分看来是提出了一种“不偏不倚理性”的后果论。
- 第173页 我们首先假定……斯坎伦所称的……WWO, 97。
- 第174页 我们应该如何理解……Rawls (TJ) 52, Nagel (1995) 182。
我不是要提出一种……“道德的最高原则”。The Groundwork, 此后记为G, 392。
参考页码指的是the Prussian Academy edition的页码，这是绝大部分英译版给出的页码。
- 第177页 人性公式……以康德的话来说：“人与一般意义上的理性存在者，是作为自在目的的存在，不是这个或那个意志可以随意用作纯粹手段的存在；相反他必须在他所有的行动中，不管行动是针对自己还是对其他理性存在者，总是被同时当作一种目的。”（G428-9）
若我想用这种诺言……G 430。
科丝嘉评论道……Korsgaard (1996) 139。
奥尼尔同样写道……见O’Neill (1989) 111。
- 第178页 科丝嘉得出……Korsgaard (1996) 140。
科丝嘉可能会说，康德的主张……欺骗总是不当的。我这里遵循Korsgaard (1996) 295-6。（科丝嘉本人并不认为欺骗总是不当的。）
- 第179页 我们现在回到康德的这个主张……康德说他欺骗的那个人“没有可能对我对待他的方式表示同意”，随后指其话语引入了他称为“其他人的原则”（G 430）。（A）是这个原则最简单的陈述。
奥尼尔同样写道……O’Neill (1989) 110。
- 第180页 给予选择的的原则……科丝嘉写道：其他人不能具有这个相同行动的目的，因为你的行动方式阻止她选择是否对该目的的实现有所贡献“（Korsgaard (1996) 138-9）。
- 第181页 我称此为同意原则。其他作者已经假定或主张这是康德的意思，例如，参

见 Hill (1992) 45。

有几个理由……不能理性地做这件事。例如，当康德主张我们不能意愿某个准则成为普遍法则时，这看来常常为真。

没有可能同意……G 429–30，我加的强调。

第182页 我们是否能够……罗尔斯表示，参见Rawls (2000) 100–91。Hill提出了一个类似的主张 (1992) 45。

第183页 罗尔斯为了支持他认为……“完全相同的法则” G 436。罗尔斯因此提出……Rawls (2000) 191 and 182–3。

这个提议贬低……提出一个新原则。C2, note on p.8。赫尔曼所称的……Herman (1993) vii。

第198页 正如我说过的，尽管自己的偏好……参见上面的66页注。(原文为000有误。)

第210页 具有这么多的资源……Metaphysics of Morals, 此后记为MM 454。参见Wood (1999) 5–8, 我从他那里获得这个以及下一个引用。

一个人可能参与……LE 179 (27: 416)。

在应用这个版本……极困难的问题，这可能是大部分富裕者所面临的最重要的道德问题。四个卓越的讨论，参见Murphy (2000), Mulgan (2001) Cullity (2004) and Pogge (2002)。参见且 (我指望) 回应: www.givingwhatwecan.org

第211页 然而，同意原则不可能……道德的最高原则。G 392。

第212页 纯粹手段原则……康德写道：“所有理性存在者都处于如下法则的支配之下：其中每一个人都要把他本人以及其他所有人，绝不能纯粹作为手段而总是要同时作为自在目的来对待” (G 433)。

第213页 卡姆拒绝……卡姆在讨论中给了我这个反驳。见Kamm (2007) 12–13 和这些页码的注释，卡姆提出了一个纯粹作为手段来对待的论说，完全不同于我的论说。基于卡姆的论说，我们是否把某人纯粹作为手段来对待，并不取决于我们对这个人的态度。即使我们没有把这个人作为手段来对待，甚至为了这个人的缘故而牺牲自己的生命，我们仍然可能是把他作为纯粹手段来对待。尽管卡姆提出了几个可行的道德主张，但我认为，她并没有描述“纯粹作为手段来对待”的日常意义。富裕者不给穷人任何东西。G 423。(康德讨论了某个“过得称心如意”的人，但她对那些穷人“没做任何贡献”。)

某个道德信念要想……康德认为，残酷地对待动物……MM 443。但是康德赞扬了莱布尼茨在显微镜下观察一个虫子后，不怕麻烦把它放回到叶子上去 (C2 5:160)。

第216页 若一个人想许假诺言……G 429。

第217页 如果我们不能诉诸 (3)……我们的主张应该只是……对于这些主张的进一步的捍卫，参见下面的42节。

第三种纯粹手段原则……例如，诺齐克做过这样的主张，见 (1974) 31。

第218页 我们接下来可以……康德主张……我们就把他们作为目的……例如，“理性

存在者……总是同时被作为目的而得到重视，也就是说，只作为必须能够自身包含在那个相同行动的目的之中的存在。”（G 429-30, 我加的强调。）

第219页 人们也许会主张，在桥梁情形中，你并不是真正地把杀我作为手段来救五个人……这个反驳是由Bennett给出的（Bennett（1995）Chapter 11）。

第224页 如果我们强制或欺骗他人……O'Neill（1989）111 and 114。

科丝嘉同样写道……Korsgaard（1996）347。科丝嘉可能只想描述康德的观点。

第225页 把某物纯粹当作手段来对待……O'Neill（1989）138。

科丝嘉也写道，基于康德的观点……Korsgaard（1996）142。

这个主张蕴含着……科丝嘉还写道……Korsgaard（1996）93。

第226页 我们会在非日常意义上……例如，罗尔斯提出，Rawls（TJ），111 and 184。罗尔斯还提出，Rawls（1999）355……这就使得……既然罗尔斯根本没有这些意义上的“正当”和“真”，我的话对他的道德理论并不构成反驳。

第227页 “纯粹作为手段”这个术语……在特殊的意义。如果某个人自杀以避免痛苦或者让自己得到性快感，康德主张这个人因此把自己纯粹作为手段来对待了。（G 429, and MM, 425。）

第229页 这个原则……按照汤普逊（Thomson）的讲法……Thomson（1990）166-168。汤普逊补充道：“然而，当数字变得极大时，有些人开始感觉到紧张，数百，数百万，整个亚洲的人口！”

第231页 我们接下来转向……例如，按照汤普逊的看法……Thomson（1990）153。汤普逊的主张是有关救四条人命的；但我认为，她会将其适用于救一条人命。

第233页 康德在对人性公式的另一个评论中写道，每个理性存在者……G 428。

伍德表示……Wood（1999）152-5。

然而，我们无需……伍德表示Wood（1999）117。

第234页 某些不当的行动……傲慢或鄙视的行动。MM 462-8。

伍德称（C）……Wood（2002）172……伍德主张，尽管（C）……Wood（1999）155。伍德指出……Wood（1999）139。最长的一本有关道德的著作中……这本书是《道德的形而上学》。

我认为，康德的论述……正如伍德本人所承认的……Wood（2006）346。康德写道，我们发展自己天赋之责。MM 444 and 392……贬低或玷污了人性……MM 423-5。纯粹的物件。MM 429-30。

第235页 伍德表示……Wood（1999）154, and 371, note 32。

当康德解释……罗尔斯称……Rawls（TJ）31, note 16. 但相对而言，伍德主张……Wood（1999）141。赫尔曼表示 Herman（1993）208, 153. 在我们考虑康德的价值主张之前……以某种方式做出回应。我这里遵循Scanlon WWO 的第1和2章。

第237页 当我们断言……Moore（1903）171。（摩尔在本段的末尾看来与这个主张相冲突。）

斯坎伦写道，理解某物的价值……WWO 99。

第238页 首先是把人的生命…… WWO 104。（原文的237页是错误的。）

第239页 斯坎伦拒绝……WWO 105。

我们接下来考虑……康德所称的尊严……价值。G 435-6。

第240页 这样的一种目的是善良意志……构成善良意愿的一种活动。赫尔曼写道，“善的领域就是理性活动和能动性（即意愿）（Herman [1993] 213）……真正天职。G 396-7。

另一种有待产生……G 433 and 438。

第三种这样的目的……相配的一切幸福。参考文献上，可参见34节开头的注释……贡献一切来产生它。C2 119康德的术语是‘das höchste Gut’，其字面意义上“最高级的善”。但康德的术语是有误导性的。正如康德自己指出的，他所称为的‘das höchste Gut’并不具有比善良意志的善性更为高级的善性，而只是最为完备的善性（C2, 111）。我的翻译是“最伟大的善”能更好地表示康德的意指，因为这种善是最伟大的，不是因为是最高级的，而是因为是最完备的。

可能还有第四种这样的目的……G 428。The Critique of Judgment 442-3。

我们可以回到……Wood（1999）133。Herman（1993）238。伍德写道：“然而，康德提出绝对命令的根据是任何具有人性（也就是出于理性来设立目的的能力，不管其意是好是坏）之存在物的价值（Wood [1999] 120-1）。康德有时言及，我们以某种方式行动不当时，我们就在抛弃我们的人性，由此，我们比纯粹物件的价值还要低。但这并不是康德的真正观点。

第241页 我一直在忽略……赫尔曼写道……Herman（1993）213。伍德甚至称……Wood（1999）121。Hill类似地写道，当康德主张个人是自在目的时，这是“内在于人的理性是这样一种目的”的简略说法（（1992）392）。

基于康德的观点……后者他称为价格。G 435……纽曼主张……Newman（1901）Vol I, 204。

第242页 接下来有人可能会主张……Hill（1992）50-57。

道德和人性……G 435。

“人性”这个词……康德确实谴责……MM 427。

第243页 赫尔曼写道，康德有关价值的主张……Herman（1993）215, 210。

康德可能不会……所称的“道德地位”。Herman（1993）129。

第244页 康德主张，最高的或最伟大的善……例如，康德写道：“这个世界的最伟大的善，至善或具有最大可能程度善一起的道德。”（LE 440 [27: 717]）（参见上述240页的注释，有关我对这个具有“最伟大”这个词之主张的翻译。）

每个人都应该努力……C2 125（当康德写“我们”时，他意指“我们所有人”）。他还写道：“这个世界里面最伟大善的产生就是由道德法则所决定的意志的必然对象。”（C2 122）“尽我们最大的能力来实现最伟大的善正是我们的义务。”（C2 143 note）

道德法则命令……C2 129。

第245页 这个理想的世界……接近康德的理想.我在此是遵循康德,他写道:“他们用这来意指这个世界里可获得的最高的善——它纵然是我们无法企及的,却是我们必须靠近的,也因而必须是通过手段的满足而予以靠近的。”(LE 253 [27:482])。他还写道:“我称之为理想的这种至善,亦即可设想的最大情形,这里任何事物都是据此决定和衡量的。在任何情况下,我们必须先构想某种典范,由此一切事物可据之得以评判。”(LE 44 [27:247])

康德主张……有些作者表示……例如Stephen Engstrom写道,基于康德的观点,幸福与德性的这种匹配会是“次优的事情”(Engstrom [1992] 769)。

康德的如下主张是不可行的……正如康德所假定的,我们常常具有这种信念……例如康德写道:“一位不偏不倚的理性观察者永远不会乐于”见到一个毫无德性的人享有幸福,当让这人的幸福被剥夺时,“每个人都赞同且觉得这本身就是善的”。他还写道:如果某人喜欢激怒和打扰爱好和平的人,他最终因他的某一次挑衅而受到厉害的鞭打……每个人会赞同它且认为这本身就是好的,即使这没有任何进一步的结果”(C2 61)。

第247页 善恶的概念绝不可……C2, 63-4。

所有的命令……G 413。

第248页 康德这里所做的似乎……“实践上必然的,也就是好的”。G 412。

K1: 善良意志……以康德的话来说,“我们根本就无法想象,这个世界上除了善良意志外,还有任何事物可以被无条件认为是善的”。G 393。他继续说这种善性是无可超越和绝对的。

康德宣称,古希腊人……C2 64。

第249页 这看来似乎是……做一切善的事情。R 72。

第250页 对道德法则的履行……LE 440-1 (27:717)。这个最高级的目的就是最伟大的善。

对道德法则最严格的遵守……R 6: 7-8 (long note)。

第251页 正如西季威克后来所论证的,这个假设也是相当可行的。ME Book IV, Chapters III to V。

第252页 摩尔接近于接受(C)……遵循某种最优的常识规则……摩尔补充道,只要这些规则既是“普遍有用也是可普遍实践的”(Moore [1903] 211-13)。摩尔否认最多的幸福就是最好;但这个与此无关。

第254页 当休谟讨论……“整体框架”……Enquiry Appendix III, 256 (我做的强调)。他还写道:“这里单个行动的结果在许多情形下都是与整个体系的结果直接相反的;前者可能是极为有害的,而后者可能是在最大程度上有利的。”在《人性论》中休谟写道:“单个的正义之举可能与公共利益或者私人利益相悖,然而,整个计划或体系肯定是极有生产力的,无论是对于社会的支持还是对于每个人的福利,实际上是绝对必要的。我们不可能把善与恶分开。”Book III, Section 2, 497 in Selby-Bigge。

第254页 当康德捍卫另一个……我会尽力产生它……SRL (8: 425-30)。

第255页 由此一个谎言……SRL 8: 426。

促进他人的幸福……LE 388 (27:651)。

如果至善存在……Metaphysik L1, 28:337 来自于大约1778年的讲座, 转引自 Guyer (2000) 94。

第257页 尽管我们可以试图……康德在别处主张……MM 385-388。我们促进自己德性的义务是我们促进自己完善这个义务中最重要的部分, 后者还包含我们作为理性存在者的其他能力。基于康德的观点……例如参见上面245页注释中的引用。

这些有关应得的主张……正如罗尔斯写道:“在CI程序中没有任何因素能够产生要求我们让幸福与德性成比例的戒律”(Rawls [2000] 316)。

理性并不赞成幸福……C1 640。他还写道:“在实践理性观念中, 还有进一步的因素伴随着对于道德法则的簪越, 即它应得的惩罚。”(C2 37)

第258页 按照决定论……康德主张……用康德的话来说,“他必须假定行动中有意志自由, 无此就没有任何道德。”(REV 8: 14)

第261页 康德称此种相容主义观点为“拙劣的托辞”。(C2 5: 96)

第262页 除非我们认……REV 8: 13。

自由的实践概念……REV 8:13。

第265页 相当于让它受控于盲目的机会 (C2 5:95)。

第266页 为了避免这种争论……引起的行动不是事件。对于这些问题的卓越讨论, 参见see Nagel (1987) Chapter VII。

第267页 人必须使得……R 6: 44。

亚里士多德类似地写道……Nicomachean Ethics 1114a19; cf.1114b30 seq。

第268页 (J) 如果我们的行动……这些行动负有责任……See Nagel (1986) Chapter 7. 我对这个论证的陈述, 部分程度上遵循了Galen Strawson, 他给出了这个论证的一个卓越版本, Strawson (1994) and Strawson (1998)。

第272页 我们可能应得许多东西……但从没有任何人应得痛苦……“没有任何人能应得痛苦”这个信念引出了很多问题, 对这些问题的讨论, 参见Sidgwick, ME, Chapter V, 特别见 section 4, and Pereboom (2001) Chapters 5 to 7。

尽管康德对……所具有的未经开发的理论能量与活力 (1993) vii。

第275页 行为的正当与否……其中有这样一些例子……C2, 第27页。C2, 第19页。G, 第423页。C2, 第34页。

一种作为“普遍允许的法则”的准则。MM, 第453页。康德还提及某种法则的普遍性, 即, 每个人都能够以某些方式行动 (G, 第422页, 强调为我所加)。

第277页 有些作者提出, 康德的意思是 (D)……例见奥尼尔 (1989), 第157页。

(奥尼尔的观点此后有所改变。例见奥尼尔 [1996], 第59页。)

但康德真正的公式是 (F)……当他讨论许假诺 (G, 第422页) 以及其他许多段落 (其中一些我在后面要提到) 时, 康德诉诸 (F)。我得承认, 康德在讨论“自杀

以免麻烦”（亦见G，第422页）的准则时没有诉诸（F）。康德没有宣称：如果我们全都接受这个准则或相信这样自杀是可允许的，那将使自杀变得不可能。这种说法显然错误。康德的主张反倒是这样的：由于这样自杀的动机关系到我们自身的福祉，而自然赋予我们这种动机是为了保存我们的生命这一目的，那就不可能存在这样一条自然法：由该自然法而来的这一动机会引导我们去自我毁灭。他写道，如果存在这样一种法则，那么自然就会陷入“自相矛盾”。这些主张没有支持康德的公式。在别处康德主张：如果每个人都“能够任意了结自己的生命，那么这样一种安排不会作为一种持续的自然秩序”。这个主张假定：如果我们都相信自杀是可允许的，人类就不可能持存，因为过多的人会在尚未成年生养孩子之前就自杀。这不太可能正确。

先考察……如赫尔曼所指出的……赫尔曼（1993），第118-119页。

第278页 似乎是足够的……赫尔曼（1993），第119页。

撒谎由于会骗人，通常……科丝嘉（1996），第136页。

每个人的财产都可被取走如果是一条普遍规则……LE，第232-233页（29:609）。

第279页 在讨论……不能“自圆其说”。C2，第19页。

上述失败……“……许假诺”。G，第402-403页、第422页。

这一法则的普遍性……G，第422页。

罗尔斯提议，在评估这一断言的过程中……罗尔斯（2000），第169页。

第281页 有鉴于此，康德或许会接受……SRL8，第425-430页。

第282页 科丝嘉……一定程度的回应……科丝嘉（1996），第95页。

第283页 奥尼尔提出了（H）的一个较弱版本……奥尼尔（1989），第133页、第135页及别处。

奥尼尔声称，这一公式谴责……奥尼尔（1989），第138-139页……“削弱了他们的能动性”……奥尼尔（1989），第215-216页。

此外，（I）……带着争胜的目标进行竞技游戏。奥尼尔（1989），第102-103页。

第284页 科丝嘉提出……科丝嘉（1996），第92-93页……使用最不拥挤的网球场……赫尔曼（1993），第138-139页……偿还信用卡上的债务……布莱克本（Blackburn）（1998），第218页……开令人意外的舞会。赫尔曼（1993），第141页。

第285页 普遍法则公式……在康德关于该公式的许多版本之中，大多数都采取命令的形式，乃至它们既不可能真也不可能假。但在第一次提出该公式时，康德写道：“我绝不当行动，除非以这样的方式行动：我能够同时意愿我的准则成为普遍法则”（G，第402页）。

康德提到……如赫尔曼所指出的……赫尔曼（1993），第123页。

就意愿某一准则……康德有时声称……例如，他写道，“准则必须这样被选择：它们将仿佛如普遍的自然法一般被持有”（G，第436页）。另见G，第421页，以及C2，第69-70页。

第286页 可允许性公式……例如，康德写道：“我究竟能否对自己说：每个人在发现自己陷入困境时都可以做出虚假的承诺？”（G，第403页），而且他提及“这样一种法则的普遍性：每个人都……能够随其所欲地许诺而又无意于守诺”（G，第422页）。类似地，在另一处康德谈到“这样一种法则：每个人都可以否认一笔无人能证明其所有者的存款”（C2，第27页）。而如我所言，康德言及一种作为“普遍允许的法则”的准则（MM，第453页）。（上述引用中的强调均为我所加。）康德公式的这一可允许性版本由斯坎伦于1983年的未刊讲稿中提出。另见博格（1998）；伍德（1999），第80页；以及赫尔曼（1993），第120-121页。

道德信念公式……康德并未明确诉诸这一公式。但据说他在讲座中曾讲道，“你应该这样行动：你的行动准则必须成为一种普遍法则，亦即，其本身必须被普遍地认可”（LE，第264页 [27：第495-496页]）。而且康德还写道：“如果每个人都……都认为，自己只要对人生厌烦至极就有权自寻短见”（C2，第69页）。（如前，强调由我所加。）

康德谈到，他所提出的……G，第403页。

第288页 如前所示，康德没有……如罗尔斯和赫尔曼所提议的……罗尔斯（2000），第167页注释3。

第289页 这一反驳可以得到部分的答复……奥尼尔（1989），第85页。

第290页 这些意见没有完全……见伍德（1999），第103-105页的讨论。

或许可以声称……罗尔斯断言……罗尔斯（2000），第187页。

康德有类似的断言……MM，第455-457页。

第291页 有些作者提出……康德经常讨论……C2，第34页。

第293页 在考察了形形色色的反驳之后……伍德（2006），第345页；以及伍德（2002），第172页。赫尔曼（1993），第104页、第132页。奥尼尔（1975），第129页、第125页。另见奥尼尔（1989），第130页。希尔（2002），第122页。

这些人认为……赫尔曼（1993），第117页。奥尼尔（1989），第86页、第98页、第103页。

第294页 用最简短但绝对可靠的方式告诉自己……G，第403页。

某人厌倦生命……G，第422页。

他仍然有足够的良知……G，第423页。

他自问，怠慢其自然禀赋的准则是否……G，第421-422页。

非常准确地确定……C2，注释8。康德还写道：“所有的义务命令都可以从这单个的命令推衍出来”，并且“这些是许多实际义务之中的几种……它们源自那一个原则，这一点显而易见”（G，第424页）。

有此罗盘在手，普通的人类理性……G，第404页。

第295页 有些作者提出……（LN2）……奥尼尔、赫尔曼、博格和卡根全都做出或讨论了此类提议。奥尼尔（1987），第87页、第130-131页；赫尔曼（1993），第147-148页；博格（2004），第56-58页；以及卡根（2002），第122-127页。

第297页 在许多情况下，要给出对某一行动的道德相关描述……以西季威克的举例

看……ME, 第202页注释。西季威克声称, 尽管这个革命者的意图是刺杀沙皇, 但是, 他并非有意杀其他人这样的说法是错误的。我认为, 还是这样说为好: 这个人在意图做的, 是以一种他知道会杀死其他许多人的方式行动。

第298页 按照我建议的康德公式的版本……这个准则不是一种策略……在康德篇幅更长的论述中, 这个准则是: “从自爱出发, 我使如下行动成为我的原则: 当活得更久的痛苦大于愉悦时缩短寿命” (G, 第422页)。这一准则可能是一种策略, 因为我们经常能够以缩短寿命的方式行动。吸烟者每当吸烟时都可能在这么做。但康德此处是在讨论单一的自杀行为。

第299页 或许有人会质疑……奥尼尔 (1975), 第112页……但如康德所指出的……G, 第424页。奥尼尔本人后来写道: “这并不是说在现实世界, 每一个欺骗者的思维之中都存在某种矛盾” (奥尼尔 [1989], 第132页)。

奥尼尔还提出……奥尼尔 (1989), 第87页……以及奥尼尔在别处所主张的……奥尼尔 (1975), 第112-117页、第124-143页; 以及奥尼尔 (1989), 第130页。在赫尔曼 (1993) 第四章与第十章中, 赫尔曼做出了类似的主张。

第307页 以自我幸福的牺牲来促进他人的幸福……MM, 第393页。

第308页 一个例子是“不生养孩子……我的这个例子取自博格 (1998), 第190页。

第309页 博格提出……博格 (1998), 第190页。在此博格是依照斯坎伦1983年的未刊讲稿。

第313页 是为一种理想事态设计的……科丝嘉 (1996), 第149页。科丝嘉做出这一主张, 并非关于康德的自爱公式, 而是关于他的人性公式。但这种区别在此是不相干的。

第314页 问题在于……希尔 (2000), 第66页。

第317页 R2也是……譬如这样的理论……据我所知, 对于规则后果论近来最佳的陈述与辩护, 见胡克尔 (Hooker) (2000)。

第318页 某一准则要通过……其他每个人都按此准则行动。我部分地依照了卡根 (2002) 及卡根 (1998) 第231-235页中的某些提议。

第319页 这个版本……拿出一大笔。见胡克尔对这个问题的讨论, 胡克尔 (2000)。

在按此方式修正……回答“大多数人不会”就足够了。如赫尔曼所指出的, 见赫尔曼 (1993), 第七章。

第321页 我希望其他每个人……MM, 第451页。我将“仁慈的” (benevolent) 换成了“有益的” (beneficent), 因为这必定是康德想表达的意思。

它不可能作为一条普遍法则……G, 第430注释。

第322页 然而, 如前所示……康德宣称, 无人能够意愿这样的法则……G, 第423页。

第325页 按照内格爾的提议……内格爾 (1970) 000; 黑尔 (1963), 第六章; 哈萨尼 (1955); 罗尔斯 (TJ) 随处可见。

第326页 现在回到康德的主张……法官可以拒绝这一抗辩。如莱布尼茨 (Leibniz)

所指出的，见莱布尼茨（1988），第56页。

按照……康德余下的反驳……然而，如康德在别处提出的……MM，第450-451页。

第328页 康德的公式……“直觉观念”……奥尼尔（1989），第94页。

第329页 对黄金规则的另一反驳……罗尔斯（TJ），第三十节。

第334页 这是个有意义的问题。如康德所指出的……例见G，第422页。

第335页 这个人如果这么讲……对此类种族主义态度和行动的谴责……见伍德（1999），第3页和第7页。

在说明……康德想象……这个人的准则不是普遍法则。例见G，第422页。

第337页 这种观点……如下情形……科丝嘉（1996），第101页。

第338页 或许有人声称……内格尔提出……内格尔（1991），第42-43页。

可能发生许多这样的情况……G，第423页（强调为我所加）。

第339页 罗尔斯对康德的公式提出……罗尔斯写道：“我相信，康德可能已假定[我们的]决定……要经受至少两种信息方面的限制。某些限制是必要的，这一点显而易见……”（罗尔斯[2000]，第175页）。在罗尔斯引证的两段之中，一段是关于目的王国公式，另一段（讨论于第176页）似乎未支持罗尔斯的解读。

根据……第三种解读……威廉斯（1968），第123-131页。

斯坎伦提出……WVO，第170-171页，以及未刊的讲稿总结。

我绝不应当行动……G，第402页。

第340页 这样一种观念……G，第432页。而且他提及“这样一种观点：每个理性存在者都必须将自身视为普遍法则的给予者……”但康德从未明确地诉诸每个人都能够合理地意愿什么。例如，刚引用的这一说法，其结尾是“通过其意志的所有准则”（G，第434页）。如果每一个人都通过他的意志准则而将自身视为法则的提出者，那他就不是在问哪些法则是每个人都能够意愿的。在其他好几处，在似乎将要诉诸每个人都能够意愿什么时，康德就回到其普遍法则公式，并建议我们诉诸自己能够意愿的法则。

第341页 用斯坎伦的话来说……WVO，第171页。

我曾说过，在考察了某些类似的反驳……伍德（1999），第172页；赫尔曼（1993），第104页、第132页；奥尼尔（1975），第125页、第129页。

第344页：然而，要用这种方式来证成某些原则……在单个回合的妥协之中达成”。

高契尔（1996），第133页……不存在任何一组原则是每个人都应当合理选择的。例见萨格登（Sugden）（1990）。

第345页 有些作者接受……杀害土著居民……某些先天残疾者的要求。高契尔（1986），第18页注释30、第268页。但高契尔拒绝诉诸……高契尔（1986），第269页。

第346页 我已拒绝……“大量显而易见的义务”。引自高契尔（1986），第17页。且如罗尔斯所言，“对于每一个人来说，依据其威慑优势并不是一种正义观”（TJ，第134页）。

罗尔斯尽管也诉诸……他所称的作为公平的正当性。罗尔斯 (TJ), 第十八-十九节。对于他这部分的观点, 罗尔斯尽管谈得很少, 但显然认为很重要。例如, 他写道: “也许我能把本书的目的最好地表达如下: 在现代道德哲学的诸多理论之中, 占支配地位的系统理论一直是某种形式的效用主义……我认为, 他们没能建立起一种可与效用主义相抗衡的、可行而系统的道德理论” (xvii)。只有在对道德的契约论论说之中, 罗尔斯才提出了一种对效用主义的替代理论。

第347页 如果一个具有慎思理性的人想要……罗尔斯 (TJ), 第569页。

然而, 既然罗尔斯……不会有充足的理由去履行正义的要求。罗尔斯 (TJ), 第575页。

我们知道人们是理性的……罗尔斯 (1996), 第49页。

第348页 由于这一界定……罗尔斯 (TJ), 第184-185页。可比较他的如下主张: “为了各方能有所选择, 可假定他们拥有对基本善的欲望”。罗尔斯 (1999), 第266页。

作为契约论者, 罗尔斯诉诸……罗尔斯就补充了一个动机假设。在诉诸其公式的过程中, 罗尔斯写道, “我们用关于合理审慎的判断 (a judgment about rational prudence) 代替了伦理的判断” (罗尔斯 [TJ], 第44页)。处在无知之幕后, 我们被 “假定彼此之间毫不关心对方的利益” (罗尔斯 [TJ], 第147页)。他还写道, 处在无知之幕后的人们 “为他们关于如下情况的合理评价所驱动: 哪种选项最可能促进他们的利益”。(1999), 第312页。罗尔斯并没有假定: 在现实世界, 每个人都是自利的。

在原则之间进行选择的过程中……罗尔斯 (TJ), 第142页。

第349页 第一, 如果每个人……将复杂得令人绝望”。罗尔斯 (TJ), 第140页。

第二, 如罗尔斯所指出的……如罗尔斯所言, “相互冷淡与无知之幕的结合达到了与仁慈同样的用处。因为这种条件的结合迫使原初状态下的每一个人都会考虑他人的利益” (罗尔斯 [TJ], 第148页)。在此罗尔斯是与不偏不倚的仁慈相比较, 且如他指出的, 无知之幕使有偏向 (partiality) 变得不可能。

罗尔斯写道, 其主要目标之一……罗尔斯 (TJ), 第22页……如罗尔斯本人所指出的……例如, 他写道, “效用主义者把适用于个人的选择原则扩展到社会” (罗尔斯 [TJ], 第28页)。

第350页 罗尔斯意识到这个问题……罗尔斯声称, 按照这个假定……罗尔斯 (TJ), 第165-166页, RE。

似乎没有任何客观的根据……罗尔斯 (TJ), 第168页。

这一评论……他写道, 他的公式……罗尔斯 (TJ), 第122、121页, RE。

第351页 旨在支持尽可能最厚的无知之幕……罗尔斯 (1999), 第335-336页。另见罗尔斯 (TJ), 第四十节。

我认为, 对无知公式的这一辩护……足以做出合理的决定。如罗尔斯在 (TJ) 第397页的主张……异己原因的決定而行动”。罗尔斯 (1999), 第265页。

第352页 继而回顾罗尔斯的断言……罗尔斯 (TJ), 第166页……就没有如他所期望的那样……提供了一种……论证。对罗尔斯论证的这一反驳, 我取自内格尔

(1973)。

第354页 决定的唯一根据……罗尔斯(TJ),第584页。

完全可能的是……罗尔斯(TJ),第29页。

第355页 对于契约观……罗尔斯(1999),第174页。

第356页 不允许以强加于少数人的牺牲……罗尔斯(TJ),第4页。

根据若干作者……尤见于内格尔(1979)中的“平等”。

第360页 斯坎伦的公式……WWO,第4-5页及第五章。

这种反驳忽略了如下事实……斯坎伦诉诸这种道义式信念之限制(尽管未用此名),见WWO,第4-5页、第194页,以及第213-216页。

第361页 ……这是不合情理的。斯坎伦(1997),第272页。

第362页 与罗尔斯一样,斯坎伦……WWO,第215页。

行为效用主义的蕴含……斯坎伦(1997),第267页。

第367页 另一种回答……诉诸……罗尔斯写道,根据直觉主义者的看法……罗尔斯(1999),第344页……根据一种罗尔斯称之为建构主义的不同观点……罗尔斯写道:

“‘接近于道德真理’这种观念在建构主义学说之中并无一席之地……也不存在任何道德事实是所采纳的原则要接近的”(1999,第353页)。

用罗尔斯的话来说……是由我们来决定的。罗尔斯(1999或论文?),第351页。

若作恰当的理解,这一方法是……斯坎伦(2003),第149页。

第370页 斯坎伦如今接受……见注释000后。

第371页 痛苦是我们有理由想要避免之事……休谟常用的“善”或“恶”……如当他写道,“除了善与恶,或者换言之,苦与乐……”《人性论》第二卷,第十九节。

严格地说,善或恶……C2,第60页。康德还主张:审慎的原则或自爱是一种假言命令,仅仅因为我们想要未来的幸福,它才适用于我们。这一主张表现为一种基于欲望的观点,而轻视我们想要未来幸福的理由。

在主张……蕴含理由的意义……在一种解读看来,斯多葛学派所做的是某种有趣的主张,即,甚至在这种非道德的意义上,痛苦也不是坏的。例见欧文(Irwin)(1996),第80页。根据某些其他作者的观点,斯多葛学派只是如康德一样在主张:痛苦在道德上并不是坏的。

第372页 罗斯也是如此……罗斯(1939),第272-284页。(罗斯做出的这些主张尽管是关于快乐,但也有意使之适用于痛苦。)

痛苦不仅对于……用内格尔的话来说……内格尔(1986),第161页……许多人认为……例如,汤姆森(Thomson)写道:“假设有人问[某事]是否一件好事。我们应该答道,“你讲的是什么意思?你的意思是‘它对于某人来说是好的吗?’”。最好有人告诉我们是否就是这个意思,或者是否意指其他某事……由此后果论就不得不退场”(汤姆森[2003],第19页)。就做出的最后断言而论,汤姆森未过早假设她的问题无法答复。

第374页 (E)每个人都应当总是做……使事情进展预期最好的任何事。这是在第

二十一节的上文所解释的意义上讲的。

第376页 有些作者主张，在运用康德的公式时……最好地满足我们作为理性行动者的真正需要……例见罗尔斯（2000），第173-176页、第232-234页。

第381页 就近原则……对这一原则的部分辩护，见卡姆（2007），第十一及十二章。

第387页 对他人的深爱……威廉斯（1981），第18页。

第389页 按照某些基于价值的客观理论……斯坎伦举出的例子……WVO，第125页。

第404页 如西季威克之所论，答案是否定的。ME，第四卷，第三至四章。

第407页 既然KC和RC的这些不同版本……使用不同意义的“应当”和“不当的”。在RP第三十七至四十三节，我讨论了其中某些问题。另见卡根（2000），以及（1998）第六章和第七章。

第410页 最高的目的……C1，A851 B879。

如果我们按这样一种方式为人处事……Metaphysik L1，28:337，引自盖耶（Guyer）（2000），第94页。

第414页 出于我先前给出的理由……见第368-370页。

第415页 接下来回顾……如斯坎伦所述，这样一种理论的一个目的在于……WVO，第11页。

第420页 情形一：某个反复无常的暴君……这些例子可能应归功于科夫卡（Kavka）（1986）。

第426页 既然Q1与Q3是不同的问题……对这些问题之间区别的一种类似的诉诸，见希罗尼米（Hieronymi）（2005）及（2006）。但希罗尼米并没有断定，绝不存在状态给予的理由。

附录B的引用部分已纳入我对该附录的注释。

参考文献

- Barry, Brian (1989) *Theories of Justice*, Volume 1 (Harvester-Wheatsheaf).
——(1995) *Justice as Impartiality* (Oxford University Press).
- Bennett, Jonathan (1974) 'The Conscience of Huckleberry Finn', *Philosophy*, Vol. 49, No 188 (April).
——(1995) *The Act Itself* (Oxford University Press).
- Blackburn, Simmon (1998) *Ruling Passions* (Oxford University Press).
- Brandt, Richard (1979) *A Theory of the Good and the Right* (Oxford University Press).
——(1992) *Morality, Utilitarianism, and Rights* (Cambridge University Press).
- Broad, C. D. (1959) *Five Types of Ethical Theory* (Littlefield, Adams, and Co).
- Broome, John (1999) *Ethics out of Economics* (Cambridge University Press).
——(即出) *Rationality Through Reasoning* (Oxford University Press).
- Cullity, Garrett (2004) *The Moral Demands of Affluence* (Oxford University Press).
- Cummiskey, David (1996) *Kantian Consequentialism* (Oxford University Press).
- Darwall, Stephen (1983) *Impartial Reason* (Cornell University Press).
——(1992) 'Internalism and Agency', *Philosophical Perspectives*, Vol. 6. Ethics.
——(1992B) Allan Gibbard, and Peter Railton 'Toward Fin de Siecle Ethics: Some Trends', *The Philosophical Review* (January).
——(1996) *Moral Discourse and Practice*, edited by Stephen Darwall, Allan Gibbard, and Peter Railton (Oxford University Press).
- Dean, Richard (2006) *The Value of Humanity in Kant's Moral Theory* (Oxford University Press).

- Engstorn, Stephen (1992) 'The Concept of the Highest Good in Kant's Moral Theory', *Philosophy and Phenomenological Research*.
- Enoch, David (2005) 'Why Idealize?', *Ethics*, 115 (July).
- Findlay, John (1961) *Values and Intentions* (George Allen and Unwin).
- Frankfurt, Harry (1988) *The Importance of What We Care About* (Cambridge University Press).
- (2004) *The Reasons of Love* (Princeton).
- Gauthier, David: MA: *Morals by Agreement* (Oxford University Press, 1986).
- (1975) 'Reason and Maximization', *Canadian Journal of Philosophy*.
- (1984) 'Deterrence, Maximization, and Rationality', *Ethics*, 94.
- (1985) 'Afterthoughts', in *The Security Gamble*, edited by Douglas Maclean (Rowman & Allanheld).
- (1997) 'Rationality and the Rational Aim' in *Reading Parfit*, edited by Johathan Dancy, (Blackwell).
- Gibbard, Allan (1965), 'Rule Utilitarianism: Merely an Illusory Alternative?', *Australasian Journal of Philosophy*, 43.
- Guyer, Paul (2000) *Kant on Freedom, Law, and Happiness* (Cambridge University Press).
- Hare, Richard (1963) *Freedom and Reason* (Oxford University Press).
- (1997) 'Could Kant Have Been a Utilitarian?', in *R. M. Hare Sorting Out Ethics* (Oxford University Press).
- Harsanyi, John (1955) 'Cardinal Utility, Individualistic Ethics, and Interpersonal Comparisons of Utility', *Journal of Political Economics*, 63.
- Herman, Barbara (1993) *The Practice of Moral Judgment* (Harvard University Press).
- Hieronymi, Pamela (2005) 'The Wrong Kind of Reason', *The Journal of Philosophy*, 102 no 9 (September).
- (2006) 'Controlling Attitudes', *Pacific Philosophical Quarterly*, 87, no1 (March).
- Hill, Thomas E. (1992) *Dignity and Practical Reason* (Cornell University Press).
- (2000) *Respect, Pluralism, and Justice* (Oxford University Press).
- (2002) *Human Welfare and Moral Worth* (Oxford University Press).
- Hooker, Bradford (2000) *Ideal Code, Real World* (Oxford University Press).
- Hume, David: *A Treatise of Human Nature*.
- Kagan, Shelly (1998) *Normative Ethics* (Westview Press).
- (2000) 'Evaluative Focal Points', in *Morality, Rules and Consequences*, edited by Brad Hooker, Elinor Mason, and Dale E. Miller (Edinburgh University Press).

- (2002) ‘Kantianism for Consequentialists’, in *Groundwork for the Metaphysics of Morals*, Immanuel Kant, edited and translated by Allen Wood (Yale University Press).
- Kahane, Guy (2004) *The Sovereignty of Suffering, Reflections on Pain’s Badness* (Oxford University D. Phil. Thesis).
- (2009) ‘Pain, Dislike, and Experience’ *Utilitas*, Vol 21, No 3.
- Kamm, Frances (1993) *Morality, Mortality, Volume One* (Oxford University Press).
- (1996) *Morality, Mortality, Volume Two* (Oxford University Press).
- (2000) ‘Famine Ethics: the Problem of Moral Distance and Singer’s Ethical Theory’ in *Singer and his Critics*, edited by D. Jamieson (Blackwell).
- (2004) ‘The new problem of distance in morality’, in *The Ethics of Assistance*, edited by Deen K. Chatterjee (Cambridge University Press).
- (2007) *Intricate Ethics* (Oxford University Press).
- Kant, Immanuel: 我的大多数引用所标的是普鲁士学院版的页码，列在大多数英文版的页边。我使用如下缩略形式：
- C1: *Critique of Pure Reason*.
- C2: *Critique of Practical Reason*.
- G: *The Groundwork of the Metaphysics of Morals*.
- LE: *Lectures on Ethics*.
- MM: *The Metaphysics of Morals*.
- R: *Religion within the Limits of Reason Alone*.
- REV: *Review of Schulz, Attempt at an Introduction*: 8: 10-14.
- SRL: *On a Supposed Right to Lie from Altruistic Motives* 8: 425-30.
- Kant: *Correspondence*, Translated and edited by Arnulf Zweig (Cambridge University Press, 1999).
- Kavka, Gregory (1986) ‘The Toxin Puzzle’, *Analysis*, 43.
- Kelly, Thomas (2003) ‘Epistemic Rationality and Instrumental Rationality: A Critique’, *Philosophy and Phenomenological Research*, Vol. LXVI, No. 3, (May).
- Kemp Smith, Norman (1915) ‘Kant’s Method of Composing the Critique of Pure Reason’, *The Philosophy Review*.
- Kerstein, Samuel (2002) *Kant’s Search for the Supreme Principle of Morality* (Cambridge University Press).
- Kolodny, Niko (2003) ‘Love as Valuing a Relationship’, *The Philosophical Review*.
- (2005) ‘Why be Rational?’, *Mind*, Volume 114.
- (2010) ‘Which Relationships Justify Partiality? The Case of Parents and Children’, *Philosophy & Public Affairs*, Volume 38 Number 1.

- (即出) ‘Which Relationships Justify Partiality? General Considerations and Problem Cases’, in *Partiality and Impartiality*, edited by Brian Feltham and John Cottingham (Oxford University Press).
- Korsgaard, Christine (1996) *Creating the Kingdom of Ends* (Cambridge University Press).
- Kuehn, Manfred (2001) *Kant* (Cambridge University Press).
- Leibniz (1988) *Political Writings* 2nd edition translated by Patrick Riley (Cambridge University Press, 1988).
- Lewis, David (1985) ‘Devil’s Bargains and the Real World’, in *The Security Gamble*, edited by Douglas Maclean (Rowman & Allanheld).
- McClennen, Edward (1988) ‘Constrained Maximization and Resolute Choice’, *Social Philosophy and Public Policy*, 5.
- Moore, G. E. (1903) *Principia Ethica* (Cambridge University Press).
- Mulgan, Timothy (2001) *The Demands of Consequentialism* (Oxford University Press).
- Murphy, Liam (2000) *Moral Demands in Nonideal Theory* (Oxford University Press).
- Nagel, Thomas (1970) *The Possibility of Altruism* (Oxford University Press).
- (1973) in his ‘Rawls on Justice’, *Philosophical Review* April, reprinted in Norman Daniels, *Reading Rawls* (Blackwell, 1975).
- (1979) *Mortal Questions* (Cambridge University Press).
- (1986) *The View from Nowhere* (Oxford University Press).
- (1991) *Equality and Partiality* (Oxford University Press).
- (1997) *The Last Word* (Oxford University Press).
- Newman, Cardinal John Henry (1901) *Certain Difficulties Felt by Anglicans in Catholic Teaching* (Longman).
- Nozick, Robert (1974) *Anarchy, State and Utopia* (Blackwell).
- (1993) *The Nature of Rationality* (Princeton).
- O’Neill, Onora (1975) *Acting on Principle* (Columbia University Press).
- (1989) *Constructions of Reason* (Cambridge University Press).
- (1996) *Towards Justice and Virtue* (Cambridge University Press).
- Parfit, Derek: RP: *Reasons and Persons* (Oxford University Press, 1984, reprinted with some corrections in 1987).
- (1986) ‘Comments’ in *Ethics* (Summer).
- (1997) ‘Reasons and Motivation’, *Proceedings of the Aristotelian Society*, Supplementary Volume.
- Pereboom, Derk (2001) *Living Without Free Will* (Cambridge University Press).

- Pogge, Thomas (1998) 'The Categorical Imperative', in *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*, edited by Paul Guyer (Rowman and Littlefield).
- (2002) *World Poverty and Human Rights* (Polity).
- (2004) 'Parfit on What's Wrong', *the Harvard Review of Philosophy* (Spring).
- Rachels, Stuart (1998), 'Counterexamples to the Transitivity of Better Than', *Australian Journal of Philosophy*, 76, No. 1 (March).
- (2000) 'Is Unpleasantness Intrinsic to Unpleasant Experiences?' *Philosophical Studies*, Vol. 99, No. 2.
- Rasdall, Hastings (1892) a review of Sidgwick's *Elements of Politics*, in *the Economic Review* 2.
- Rawls, John: TJ: *A Theory of Justice* (Harvard University Press, 1971).
- (1996) *Political Liberalism* (Columbia University Press).
- (2000) *Lectures on the History of Moral Philosophy*, edited by Barbara Herman (Harvard University Press).
- (2001) *Justice as Fairness* (Harvard University Press).
- (2001B) *Collected Papers*, edited by Samuel Freeman (Harvard University Press).
- Raz, Joseph (2000) *Engaging Reason* (Oxford University Press).
- Regan, Donald (1980) *Utilitarianism and Cooperation* (Oxford University Press).
- Reid, Thomas (1983) *The Works of Thomas Reid* (Georg Olms Verlag).
- Ridge, Michael (2006) 'Introducing Variable Rate Rule-Utilitarianism', *The Philosophical Quarterly* (April).
- Ross, Sir David (1930) *The Right and Good* (Oxford University Press).
- (1939) *Foundations of Ethics* (Oxford University Press).
- Ruskin, John (1903) *The Works*, edited by E. T. Cook and Alexander Wedderburn, Volume xi (London).
- Scanlon, T. M. WWO: *What We Owe to Each Other* (Harvard University Press, 1998).
- (1997) 'Contractualism and Utilitarianism', *Moral Discourse and Practice*, edited by Stephen Darwall, Allan Gibbard and Peter Railton (Oxford University Press).
- (2002) 'Replies', in *Social Theory and Practice* Vol. 28.
- (2003) 'Rawls on Justification', in *The Cambridge Companion to Rawls*, edited by Samuel Freeman (Cambridge University Press).
- (2003B) 'Value, Desire, and the Quality of Life', in *The Difficulty of Tolerance* (Cambridge University Press).

- (2007) ‘Structural Rationality’, in *Common Minds*, edited by Geoffrey Brennan, Robert Goodin, and Michael Smith (Oxford University Press).
- (即出) *Being Realistic about Reasons* (Oxford University Press).
- Schneewind, Jerome (1977) *Sidgwick’s Ethics and Victorian Moral Philosophy* (Oxford University Press).
- Schultz, Bart (2004) *Henry Sidgwick: Eye of the Universe, an Intellectual Biography* (Cambridge University Press).
- Sidgwick, Henry: ME: *The Methods of Ethics* (Macmillan and Hackett various dates).
- HSM: *Henry Sidgwick: A Memoir*, by A. S. and E. M.S (Macmillan).
- (2000) *Essays on Ethics and Method*, edited by Marcus Singer (Oxford University Press).
- Smith, Michael (1994) *The Moral Problem* (Blackwell).
- (2004) *Ethics and the A Priori* (Cambridge University Press).
- Stratton-Lake, Philip (2004) *On What We Owe To Each Other* (Blackwell).
- Strawson, Galen (1994) ‘The Impossibility of Moral Responsibility’, *Philosophical Studies* 75.
- (1998) ‘Free Will’ in *the Routledge Encyclopaedia of Philosophy*, edited by E. Craig (Routledge).
- Sugden, Robert (1990) ‘Contractarianism and Norms’, *Ethics* 100.
- Suikannen, Jussi (2009) *Essays on Derek Parfit’s On What Matters*, edited by Jussi Suikannen and John Cottingham (Blackwell).
- Temkin, Larry (1987) ‘Intransitivity and the Mere Addition Paradox’, *Philosophy & Public Affairs*, 16, no. 2.
- (即出) *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford University Press).
- Thomson, Judith (1990) *The Realm of Rights* (Harvard University Press).
- (2003) *Goodness and Advice* (Princeton University Press).
- Williams, Bernard (1981) *Moral Luck* (Cambridge University Press).
- (1985) *Ethics and the Limits of Philosophy* (Fontana).
- (1995) ‘Internal Reasons and the Obscurity of Blame’, in his *Making Sense of Humanity* (Cambridge University Press).
- (1995B) *Making Sense of Humanity* (Cambridge University Press).
- (2003) *The Sense of the Past* (Princeton University Press).
- (2006) *Philosophy as Humanistic Discipline* (Princeton University Press).
- Williams, T. C. (1968) *The Concept of the Categorical Imperative* (Oxford University Press).

Wodehouse, P. G. (1952) *Pigs Have Wings* (Ballentine Books).

Wood, Allen (1999) *Kant's Ethical Thought* (Cambridge University Press).

—(2002) *Groundwork for the Metaphysics of Morals*, translated by Allen Wood (Yale University Press).

—(2006) 'The Supreme Principle of Morality', in *The Cambridge Companion to Kant and Modern Philosophy*, edited by Paul Guyer (Cambridge University Press).

索引

我在 1 到 28 页已经给出了主要主张与论证的内容提要，这里就不再重复其中的某些信息。这个索引给出的页码主要是为了：（1）我对各种主题的主要论证，小于 28 页的参考的是内容提要；（2）我在其他地方对这些主题的零散评论；（3）我对其他主题简要评论；（4）某些其他人的主张。有些条目出现了重叠，要么是因为这些主题是重叠的，要么是为了减少条目数，只是告诉你去参考其他的条目。

Act Utilitarianism（行动效用主义）：190；其享乐主义形式，246，更好地重述为有关痛苦之主张，463；作为不偏不倚理由之行动后果论的一种形式，它也许可以称为是道德的外部对手，168-9，500

Act-and-Rule Consequentialism（行动暨规则后果论）：251-6

Actualism about the goodness of events（有关事件善性的实现主义）：236

aesthetics, reasons or causes（审美理由或原因）：53-4

行动的不当性是否能取决于行动者的准则：参见 Kant's Formulas（康德的各种公式），the Mixed Maxims Objection agent-causation and free will（对行动者因果性和自由意志的混合型准则反驳）：266-7

agony（痛苦），参见 pain（疼痛）

Agony Argument（痛苦论证）：73-82，456-7

Aid Agency（援助机构情形），以及同意原则是否要求过高：209-10

aim-based reasons（基于目标的理由）：58-60；可能是派生的，66（“第五种……”）

All or None Argument（全部或全无论证）：83-91

Analytical Subjectivism（分析性主观主义）：70-3，75，125；以及卷二

anti-Utilitarian protective aim（反效用主义的保护性目标）：356-7

Anxiety and Mistrust Argument（焦虑和不信任论证）：363-5；and Scanlon's Individualist Restriction（与斯坎伦的个人主义限制），473

apparent reasons (显见理由): 35, 454 (on 35), 111; 以及参见 Reasons and Rationality (理由与合理性)

arbitrary distinctions and procedural rationality (任意区分和程序合理性): 56-7, 79-80

Aristotle (亚里士多德), 论形成我们自己的性格 : 267

attempted murder (未遂谋杀), 以及责备是否可以取决于运气 : 155-8, 461-2

Self-Defence (自卫), 以及若伤害某人作为手段, 我们是否把他纯粹作为手段对待 : 221-2, 225

Audi, Robert (罗伯特·奥笛): xlvi

Autonomy (自主性): 对自主性的尊重, 66 (“第六种……”); 自主性和生命的价值, 239; Kant’s Formula of Autonomy (康德的自主性公式), 340; 罗尔斯对其无知之幕的康德式诠释, 351

Bad Samaritan(不乐善好施者), 把人们不是纯粹当作手段, 而是纯粹当作物件来对待 : 226-7

bargaining power and Hobbesian Contractualism (议价能力和霍布斯式契约论): 345, 349, 357

belief-relative senses of ‘ought’ and ‘wrong’, and some other senses (“应当” 和 “不当的” 的信念相对意义以及其他意义): 150-62

beliefs and desires (信念和欲望), 大多数都是非自愿的 : 47-50, 420-32

Bentham, Jeremy (杰里米·边沁), 论对未遂谋杀的惩罚以及威慑的经济性 : 461-2

Berker, Selim (瑟力姆·贝克): xlvi

best for someone (对……是最好的): 其日常的广泛含义和福祉的不同理论, 39-40; 蕴含理由的意义, 41; 基于当前选择的意义, 103, 496-7, 它具有说不通的含义, 104; 享乐主义的意义和时间中立的基于欲望的意义, 它们就其自身来使用的话, 只是陈述了隐藏的同义反复, 105; 若把它们与蕴含理由的意义组合起来我们就能陈述实质性的主张, 106; 主观主义者做不到这一点, 107, 102

bias towards the near (偏向眼前), 46, 57; towards the future (偏向未来), 197

Blameworthiness (可责备性): 可责备性与 “不当的” 的事实相对的、信念相对的和道德信念相对的意义, 153-8; 如准康德主义者所主张而康德主义者所否定的, 可责备性是否能部分地取决于运气, 154-8, 461-2; “我们尽力避免可责备性” 的理由, 174; 可责备性与应得痛苦不是一样的, 272

blameworthiness sense of ‘wrong’ (“不当” 的可责备性意义): 165, 168-71

Blind man crossing the street, and ought implies can (盲人过街和应当蕴含能够): 260-1, 108

Blue’s Choice, and Rawls’s thin theory of the good (布鲁的选择和罗尔斯稀薄的善理论): 104-6

Bomb, and the priority of our duty not to kill (炸弹情形和我们不杀人义务的优先性): 449

Brandt, Richard (理查德·布兰德特): 126

Brewer, Talbot (塔尔波特·布鲁尔): xlvi

Bridge (桥梁情形): 作为手段的伤害和纯粹作为手段的对待: 218–220, 223–4, 228–30; 桥梁情形和斯坎伦式契约论, 364–6; and the Wrong-Making Features Objections to the Kantian Argument (桥梁情形与对康德式论证的致不当特性反驳), 393–7; and deontic reasons (桥梁情形和道义式理由), 450–1

Broad, C. D. (C. D. 布劳德): xxxix

Broome, John (约翰布罗姆): xlvi; 布罗姆论公平和给予人们得到利益的机会, 463; 布罗姆理由、规范性和合理要求, 参见 Broome (即出)

buck-passing view of goodness, Scanlon's (斯坎伦的善性的张冠李戴观): 39, 495

Chang, Ruth (张露丝): xlvi

Chislenko, Eugene (尤金·基什连科): xlvi

choice-based reasons (基于选择的理由): 64–5, 90; choice-based sense of 'best for someone' (“对某人是最好的”之基于选择的意义), 103–5, 496–7

Choice-Giving Principle (给予选择的原则): 180–1, 192

codes of honour, normativity and reasons (荣誉规则、规范性和理由): 144–8

Coercion (强制): and the impossibility of consent (强制和同意的不可能性), 177–9; and actual consent (强制和实际的同意), 191–200; and treating as a mere means (强制和纯粹作为手段的对待), 224–6; and Kant's Impossibility Formula (强制和康德的不可能性公式), 283

Cohen, Jerry (杰里·科恩): xlvi

collectively self-defeating principles (集体式自我挫败原则): 306–7, 468–9

common sense morality (常识道德): 149; and Sidgwick's Methods (常识道德与西季威克的《方法》), xl; and promoting happiness (常识道德与促进幸福), 251–2; 常识道德在许多群已困境中直接地集体式自我挫败, 304–7; 常识道德会要求联合的有条件诺言, 468; 但常识道德与最优原则有重叠, 因为行动后果论是间接地集体式自我挫败的, 405–6, and Parfit, RP, Part One (帕菲特《理由与人》第一部分)

concealed tautologies (隐藏的同义反复): 定义性同义反复, 70–2; 同义反复被误认为是实质性主张, 72–3; Analytical Subjectivism (分析性主观主义), 72, 75; 罗尔斯的稀薄善理论和其他此类理论, 104–6; 合理地行动和我们期望效用的最大化, 125; 罗尔斯提出的“正确”与“真”之意义, 226–7; Moore's sense of 'right' (摩尔有关“正当的”之意义), 247, 464–5; 斯坎伦观点的一种形式, 369–70; 以及卷二。

concepts, normative (规范概念): 非道德的规范概念, 31–42, 454; 道德的规范概念,

Consent Principle (同意原则): 8-10, 177-211; 康德有关同意的主张, 177; 给予选择的原则, 180; 同意原则, 180-1; 康德的理想, 182; 罗尔斯的诠释和康德观念的活力, 182-3; 我们是否能实现康德的理想, 191, 207, 210-11; 实际同意在道德上的重要性; **the Veto Principle (否决权原则)**, 192; **the Rights Principle (权利原则)**, 194-8; **assumed hypothetical present consent, past actual consent, and retroactive endorsement (当前认定的假设性同意、过去的实际同意和回溯性认可)**, 195-8; 这些时间差别的意义, 198-200; 我们本可理性地同意的不当行动, 200-4; 同意原则是否要求我们行不当之举, 203-7; 同意原则是否要求过高, 207-11; 同意原则是否总是谴责未经人们同意且以伤害他们作为手段的行动, 218-21, 230; 同意原则在某种意义上比康德的普遍法则公式要更好, 329, 333, 336-7; 但它不可能是道德的最高原则, 211, 200-1; **the Formula of Universally Willed Acts (普遍意愿行动之公式)**, 483-4

Consequentialism (后果论): 22-6, 373; **common-aim and value-based versions (共同目标和基于价值的后果论版本)**, 246; **value-based versions use the impartial-reason-implying sense of 'best' (使用蕴含不偏不倚理由意义上之最好且基于价值的后果论版本)**, 41-2; 非后果论者也能使用这种意义上的“最好”, 401; 效用主义, 373; 后果论有关何物最好之主张不限于结果或将来: 373-4; 正当与不当行动之好坏, 473-5; 有关终极重要事物之信念, 417; 后果论准则可以应用于行动之外的事物, 373-5; 直接和间接形式的后果论, 374-5, 475:

Act Consequentialism, or AC (行动后果论或 AC): 246; 摩尔无足轻重的分析性行动后果论版本, 247, 464-5; 如果我们都接受或遵循 AC, 事情是否会进展最好, 404-6, 482-3; 如果不会, 那么 AC 将是间接地集体式自我挫败的, 会是我抹消的, xxxviii (因此这个效用主义结论…), 和帕菲特的《理由与人》, 第 1 章; 行动后果论和群己困境: 305-7; 行动后果论和康德有关撒谎的主张, 254-6; **Impartial-Reason Act Consequentialism as an alternative to morality (不偏不倚理由的行动后果论是取代道德的)**, 168-9, 143-4, 498 (on 133), 500 (on 169) (这里指的是 500 页里那个针对 169 页的注释, 下同); 行动后果论和行动效用主义, 190; 行动后果论和斯坎伦契约论, 362-4, 473 (on 363); **Hedonistic Act Utilitarianism (享乐主义行动效用主义)**, 269, 169

Motive Consequentialism (动机后果论): 动机后果论和系统性的德性伦理学, 375; 动机后果论作为更宽泛理论之一部分, 407; 动机后果论的个体主义版本, 475

Maxim Consequentialism and Kant's Formula of Universal Law (准则后果论和康德的普遍法则公式): 375-7

Rule Consequentialism, or RC (规则后果论或 RC): 不是由于行动使得事情进展最好, 而是由于行动为最好的或最优规则或原则所要求或允许, 375; 规

则后果论与行动后果论之间的关系，部分地是一种历史性探讨，251-6; the universal acceptance version of RC (RC 的普遍接受版)，374-7; the universal compliance version (RC 的普遍服从版)，405-7; 哪种更接近行动后果论，406-7, 482-3; and the Ideal World Objections (理想世界反驳)，316-20; 为应对这个反驳而做的修正，319, 469; 当修正时更为接近 AC, 319; 这个修正引出的某些问题放到后面讨论，341, 479; Rule Utilitarianism and Rawlsian Contractualism (规则后果论和罗尔斯式契约论)，349-55; Kantian Rule Consequentialism (康德式规则后果论)，411, 484-5; the Kantian Argument for (规则后果论的康德主义论证)，377-411; 作为三重理论的一部分，411-8; 规则后果论和康德的主张，249-51, 254-57, 408-10; 以及参见卷二

Contractualism (契约论) : 20-26; 342-370, 377-419, 471-3, 484-5

Hobbesian Contractualism (霍布斯式契约论) : the Rational Agreement Formula (理性协议公式)，343-4; 具有某种吸引力，但面临两个反驳，344-5; Gauthier's minimal morality (高契尔的最低度道德)，346; 高契尔主张，只有他的理论才能表明不当地行动不可能是合理的，但这个论证看来是失败的，433-447, 485-92

Rawlsian Contractualism (罗尔斯式契约论) : 346-58, 367; rightness as fairness (作为公平的正当性)，Rawls's contractualist moral theory (罗尔斯的契约论道德理论)，346, 511; 罗尔斯有关合理性的假设，347-8; 罗尔斯在理性协议公式上添加了无知之幕，348-9; 罗尔斯是否成功地提出了取代所有形式效用主义的一种选项，349-55; the Equal Chance and No Knowledge Formulas (机率均等公式和无知公式)，350-1; the Maximin Argument (最大的最小值论证)，352-4; 罗尔斯式契约论与康德式契约论的比较，356-8; 进一步的评论，472-3

Scanlonian Contractualism (斯坎伦式契约论) : Scanlon's Formula (斯坎伦的公式)，360; Scanlon's intended senses of 'reasonable' and 'unreasonable' (斯坎伦意图“合情理”和“不合情理”具有的那种意义) : 360, 473; Scanlon's appeal to the Deontic Beliefs Restriction (斯坎伦诉诸道义式信念限制)，360-363; 对这种限制的一种辩护，366-70; 康德式契约论和斯坎伦式契约论给出的论说，为什么应该不是关于不当性本身的论说，而是关于致使行动不当的一种更高层次属性的论说，368-70; 我们应该不是在应用这些契约论公式时，而是在决定是否接受这些公式时，才诉诸我们的道德信念，370; 斯坎伦式契约论的修正版如何才能与康德式契约论和规则后果论相组合，411-16; 以及参见卷二

Kantian Contractualism (康德式契约论) : 康德的普遍法则公式应如何修正，从而它能变成康德式契约论公式：14-17, 19-20; the Mixed Maxims Objection (混合型准则反驳)，290-3, 以及最好的回应，293-300; the Non-

- Reversibility Objection (不可逆反驳), 334–8, 以及最好的回应, 338–42, 470–1; 康德式契约论、霍布斯式契约论和罗尔斯式契约论的比较, 355–60; and the Deontic Beliefs Restriction (康德式契约论和道义式信念限制), 366, 后者是可以证成的, 367–70
- Kantian Contractualist Argument for Rule Consequentialism (对规则后果论的康德式契约论的论证): 23–26, 377–403; 是否每个人都能理性地选择最优原则, 380–399; self-interested reasons and the High Stakes Objection (自利理由与高筹码反驳), 380–4; altruistic reasons, (利他理由) 385–6, 387–9; deontic reasons and the Deontic Beliefs Restriction (道义式理由和道义式信念限制), 386–7; the Wrong-Making Features Objection (致不当特性反驳), 389–99, 448–51; 是否每个人都能理性地选择任何有意义且非最优的原则, 399–400; 康德式公式蕴含着规则后果论, 400–1, 正如我们本应预料到的, 401–2, 这是康德的普遍法则公式唯一能成功的方式, 402–3; 对康德主义论证的进一步反驳, 476–481; 以及卷二
- Kantian Rule Consequentialism (康德式规则后果论): 它能采取其他的形式, 404–8; 行动后果论准则是否能被意愿为普遍法则, 482–3 (两个注释都是 406 页的); 康德式规则后果论与康德自己的观点, 340, 408–410, 也见 249–51, 254–7; 康德式契约论和斯坎伦式契约能够组合, 对于三重理论的趋同论证: 25–6, 411–16; 攀登山峰, 417–19; 以及参见卷二
- contradiction-in-conception test and the Impossibility Formula (观念矛盾的检验和不可能性公式): 465 (on 275)
- contrary to duty or in conformity with duty, the aims of Kant's Formula of Universal Law (违背或符合义务, 康德普遍法则公式的目标): 293–4
- contrary to reason, irrational or open to rational criticism (违背理性, 不理性或受到理性批评): 56–7, 119–25
- Convergence Argument for the Triple Theory (三重理论的趋同论证): 411–16
- 'could', categorical and hypothetical senses of (绝对意义或假设意义的“能够”): 260–3
- criterion of wrongness (不当性的标准): 190–1; 康德的公式是否想以及能否提供这样一种标准, 293–4; and the Triple Theory (三重理论), 414–15
- Cullity, Garrett (加内特·卡内提): xlvi
- cyclical preferences (循环偏好): 127–8
- Dancy, Jonathan (乔纳森·丹西): xlvi
- Darwall, Stephen (斯蒂芬·达沃尔): on Metaphysical Naturalism, normativity, and motivating force (论形而上学自然主义、规范性和驱动力), 110, 497
- Deception (欺骗): 当欺骗不当时, 不是因为它使得同意不可能, 177–9; 可能也不涉

及把他人纯粹当作手段来对待, 224-6; 康德有关欺骗的主张, 234-5, 254-5; 康德的不可能性公式并不谴责欺骗, 278

decisive reasons, and the decisive reason-implying senses of 'should', 'ought', and 'must' (决定性理由以及蕴含决定性理由意义上的“应该”、“应当”和“必须”): 32-3, 454 (on 33)

decisive-moral-reason senses of 'ought' and 'wrong' (决定性道德理由意义上的“应当”和“不当”): 它不同于道德上的决定性理由, 167; 这些意义蕴含着某个行动是否不当没有任何实践影响, 166-7, 172-3; 这些意义应用于某些行动时是最说不通的, 448-51

Deliberative Theory (慎思理论): 理由的慎思理论或慎思主观主义: 62-4, 78, 95; 合理性的慎思理论, 78-81, 103-4, 347; of well-being (福祉的慎思理论), 103-4, 496-7

demandingness of morality(道德的过高要求): 148-9; 道德的过高要求与我们富裕者, 207-211, 319, 501 (on 210)

deontic reasons, reasons given by an act's wrongness (道义式理由, 由行动的不当性给予的理由): 它不同于道德理由, 166-7; 道义式理由与不当性的类型, 172-4, 448-51; 道义式理由与同意原则, 201-6

Deontic Beliefs Restriction (道义式信念限制): 它与康德的公式, 287-8; 它与契约论公式, 360-2, 416; 它的辩护, 366-70; 它与对规则后果论的康德主义论证, 386-7; 它与致不当特征反驳: 390-8, 448-51; 它与道义式价值限制, 473-4

depression, and justified coercion (抑郁情形和有证成的强制) 197-8

derivative reasons (派生理由): 39; and goodness or badness (它与好处和坏处), 42; and subjective theories (它与主观理论), 66

Desert (应得): 13-14; 许多人相信幸福与痛苦可以是应得的, 142; Kant's ideal world (康德的理想世界), 240; Kant's Formula of the Greatest Good (康德最伟大善的公式), 245; 康德的其他主张并不支持其有关应得的信念, 257; 应得痛苦意义的道德责任, 264; 康德认定, 如果我们的行动纯粹是时间中的事件, 我们就不可能应得痛苦, 263-6; 对此观点的一种论证, 267-8; 既然我们的行动纯粹是时间中的事件, 我们就不可能应得痛苦, 269-72; 对康德理想世界的重述, 409-10

desire-based or aim-based reasons (基于欲望或目标的理由): 参见 Reasons (理由), Subjectivism (主观主义)

desire-dependent or aim-dependent reason-giving facts (依赖于欲望或目标且给予理由的事实): 67-9

desire-based theories of well-being (基于欲望的福祉理论): 40, 105-6

Desires (欲望): 狭义和广义以及心理利己主义, 43; 目的式欲望和工具性欲望, 44; 使某个欲望成为目标, 44; 我们的欲望是否能赋予我们想要之物以价值: 46, 55, 67-9; 元享乐式欲望和享乐式喜恶之差别, 54-6

Desperate Plight (绝望困境情形), 把某人纯粹当作手段来对待: 225

Determinism (决定论), 258; 被认为会剥夺掉我们道德所要求的那种自由, 258–260; 与“应当蕴含能够”的那种“能够”之意义是相容的。260–3; 有时候被混淆为宿命论, 261–2; 如果我们的行动是彻底地因果决定的, 我们是否能应得痛苦, 264–72, 506 (on 268 and 272); 康德接近于回应不能, 271; 如果某个事件部分地不是被引起的, 而是随机的, 这也不会有影响, 265; 是否存在第三种选项, 266–9

dignity, Kantian (康德式尊严): 239–44

Direct Consequentialism (直接后果论): 374–5

Discount rate with respect to time, 时间的折扣率看来是说得通的: 459 (on 121)

distributive justice (分配正义): and the Golden Rule (分配正义与黄金规则), 330; 分配正义与罗尔斯式契约论), 353; 西季威克几乎完全忽略了分配正义, 460 (on 133)

divine command sense of ‘wrong’ (“不当”的神令论意义): 166, 170, 173

Drowning Child (溺水的孩子): 利己主义者救了他并将其纯粹当作工具, 但其行动并无不当, 216, 正如第三种纯粹手段原则所承认的, 217; 康德的普遍法则公式谴责这种行动为不当的, 290–1

Dualism of Duty and Self-Interest (义务与自我利益的二元论): 141–9

Dualism of Practical Reason (实践理性的二元论): 130–144, 498–9 (on 133 and 143)

duties of justice (正义之责) (其履行可以不管动机) 和德性之责 (其履行不能不管动机): 290–1

each-we dilemmas (群己困境): 17; 这些包括所谓的囚徒困境, 301–3; “重复的囚徒困境”如何不是真正的群己困境, 467; 有许多真正的困境, 大部分涉及许多人, 303–4; 政治的、心理的和道德的解决方案, 304–6; 常识道德与联合的有条件协议, 468; 康德的自然法则公式若被接受就能提供最好的解决方案, 305–7; 正如康德的建议, 306–7; 应用于其坏效果分散特别广时尤其如此, 307–8

Earthquakes (地震情形): First Earthquake (第一种地震情形), 作为对同意原则的检验, 185–6, 192–3, 201–2, 208; Second or Means, (第二种地震情形或手段情形) 用来阐明给予同意的道义式理由, 201–4; Third and Fourth (第三种和第四种地震情形) 作为纯粹手段原则的检验, 222–3, 231–2

Egoism (利己主义): 心理利己主义, 43; Rational Egoism (理性利己主义): 52; 有些版本是隐藏的同义反复, 125, 135; 实质的理性利己主义不是一种道德观点, 而是道德的外部对手, 166; 利己主义的卑鄙性, 169; 但不可否认是说得通的, 498 (on 133), 499 (on 143); 利己主义和非利己主义理由, 130–44, 358–9; 利己主义和群己困境, 306–7; 利己主义和慎思主观理论, 347–8; 作为对康德式论证的反驳的高筹码利己主义, 383–4; 利己主义和高契尔的理论, 433–47

Egoist, my imagined (虚构的利己主义者): 把别人纯粹当作手段来对待以有益于己,

但并未行动不当, 216, 正如第三种纯粹手段原则所承认的, 217, 但康德的普遍法则公式谴责这样的行动, 290-1

Egoistic Maxim (利己主义准则): 利己主义准则和混合型 准则反驳, 290-3; 利己主义准则可以被某些行为不当者意愿为普遍的, 330-1

Einstein (爱因斯坦): xl, 329

Emergency Principle (紧急原则): 365-6

empirical beliefs (经验信念): 128-9

ends-in-themselves (自在目的): 239-40, 464, 501, 504

ends-to-be-produced (有待产生的目的): 239-40

Enoch, David (大卫·伊诺克): xlvi

epistemic reasons (认知理由): 持有信念的真值相关理由, 参见 Reasons (理由)

Error-Free Desire Theory (无错的欲望理论): 60, 94-5

Equal Chance Formula, and Rawlsian Contractualism (机率均等公式和罗尔斯式契约论): 350-1

equal chances, giving people, and Expectabilism (给予人们均等的机率和期望主义), 462-3

Equal Claims (平等的要求权), 我们是否能理性地同意只给予得救的均等机率: 206-7

Euthanasia (安乐死): 200, 204-5, 238-9

events, in the wide sense (广义上的事件): 43-4, 236

evidence-relative senses of 'ought' and 'wrong', and some other senses (‘应当’和‘不当’的证据相对的意义和其他意义: 150-62; ‘实践上应当’的证据相关的意义, 163

existent ends (存在性目的): 239-40

Expectabilism, expectably-best (期望主义, 期望上最好的): 尽力做在事实相对意义上是正当的事情, 159-62, 462-3; 期望主义与没必要的分歧, 171, 374-5

extrinsic goodness (外在善性): 52, 455

fact-relative senses of 'ought' and 'wrong' (‘应当’和‘不当’的证据相对的意义): 和其他意义, 150-62, 374

False Belief (虚假信念), 耶和華见证会教徒能够有效地拒绝同意: 198

Fatal Belief, and justified deceit (致命的信念和有证成的欺骗): 178

fatalism and determinism (宿命论和决定论): 261-2

Fire, and deontic reasons to give consent (火灾情形与给予同意的道义式理由): 204-5

Fitzpatrick, William (威廉·菲兹帕特里克): xlvi

following a principle or rule (遵循原则或规则): 405-6

Frankfurt, Harry (哈里·法兰克福): 论作为重要的或对我们重要的, 96–101

free riders (搭便车者): 305

free will of the kind required for morality (道德所要求的那种自由意志), 258–263;
为了应得痛苦所要求的那种自由意志, 263–72

future of humanity (人类的未来): 31, 419

Future Tuesday Indifference (对未来周二的漠不关心): 56, 79, 120–4

Gangster, and treating people merely as a means (土匪情形, 以及把人们纯粹当作手段来对待): 216, 228, 231–2

Gauthier, David (大卫·高契尔): 契约论理论, 344–6; 高契尔有关合理性的主张 433–447, 485–92

Gibbard, Allan (阿兰·吉巴德): 61, 483

global warming (全球变暖): 307, 374, 419

God (上帝): 上帝与道德, 166, 466 (on 286); 上帝与我们不行动不当的理由, 173;
上帝与康德的本体世界, 269; 上帝与是否有任何人能应得痛苦, 270–2

Golden Rule (黄金规则): 人类最早也是最广为接受的道德观念, 321, 469–70; 康德的批判是失败的, 321–7; 黄金规则的优缺点, 328–30; 比康德的公式更好地实现不偏不倚, 334–8

Good (and bad) (好与坏): 涉及理由的概念和属性, 38–42, 45–7, 235–7;

good for (对……是好的): 涉及理由的意义, 39–41, 102; 基于当前选择的意义, 103, 496–7; 其他的意义, 105–7; 参见 best for someone (对某人是最好的)

good, period, or impersonally good in the impartial-reason-implying sense (好的, 勿用多言, 或在蕴含不偏不倚理由的意义上是非个人地好的): 41–2, 371–3, 235–9; 事件可以因为它们对于人们是好的而非个人地好的 42; 有关什么是好的之不同观点, 52, 130, 373–4; 道德上的善性与恶性, 473–4; 最好的和期望上最好的, 159–60; “最好的”之蕴含不偏不倚理由的意义与道德的不偏不倚观察者论说之间没有任何联系, 458 (on 101); 以及参见 Reasons (理由)

主观主义者不能使用这些涉及理由的概念: 46, 93–5, 101–2, 105–7

intrinsic goodness (内在善性): 38–9, 42; 工具性善性, 50; 外在善性, 52, 455;
内在善性和欲望的合理性), 129, 497 (on 107); 事件和事情的善性, 236–7
要被尊重而不是促进的价值, 237–9; “正当”之促进善的意义: 247

Greater Burden Principle and Scanlonian Contractualism (更大负担原则和斯坎伦式契约论): 361–2

Greatest Good (最伟大的善): 参见 Kant's Formulas (康德的各种公式)

Happiness(幸福): 幸福和福祉, 39; 康德有关幸福的主张, 243, 410; 幸福和效用主义, 246; 幸福和痛苦, 463 (on 169); 幸福与人的分立, 133; 西季威克对规范享乐主义

的论证, 453 (on xl); 我们若都有行动后果论者的动机对幸福的影响, 404, 387 (“深刻情感……”); 以及参见 pain (疼痛, 痛苦)

harming people as a means (以伤害别人作为手段): 219; 可能没有涉及把这些人纯粹作为手段来对待 221-8; the Harmful Means Principle (有害手段原则), 229, 361-6, 464; 与斯坎伦式契约论) 360-6

Harsanyi, John (约翰·海萨尼): 326

Hedonic 享乐的: 涉及苦与乐、痛苦与幸福; 享乐式理由, 52-6; 当前感知的享乐式喜恶, 极为不同于元享乐式欲望, 52-7, 67-8; 参见 pain, happiness(疼痛、痛苦, 幸福)

Hedonistic(享乐主义的): 人或理论极为看重享乐式理由、价值或欲望; 心理享乐主义, 44; 福祉的享乐主义理论, 39-40, 还原分析性版本), 105; 享乐主义的行动效用主义: 246, 169; 康德早期有关享乐主义的规则效用主义的主张, 255-6, 410

Herman, Barbara (芭芭拉·赫尔曼): 论康德的伟大, 183; 康德的伦理学是否最好理解为价值伦理学, 235, 240-3, 248, 503; 论康德的不可能性公式, 277-8; 论康德的普遍法则公式, 285, 288, 293; 以及卷二

High Price and the Consent Principle (高价格情形和同意原则): 207

High Stakes Egoism as an objection to the Kantian Argument (作为对康德式论证反驳的高筹码利己主义): 383-4; 类似的反驳涉及到救那些我们所爱的人, 388-9

High Stakes Objection to the Kantian Formulas (对康德式公式的高筹码反驳): 331-3

higher-level wrong-making properties or principles(更高层次的致不当属性或原则): 369, 414-15, 476

Highest Good (最高的善), 参见康德的最伟大善公式(如 503 页末尾的解释)

Hill, Thomas (托马斯·希尔): 论合理性的价值, 242; 论康德的普遍法则公式, 293; 论理想世界反驳, 314; 论同意原则, 501 (on 181)

Humanity(人性): 参见 Kant's Formulas(康德的各种公式); 人性在康德那里的意思: 234-5, 241-2, 255, 464 (on 24)

Hume (休谟): xxxiii, xl, xlv; 论持有欲望的理由, 100, 113, 118, 371, 以及卷二; 论单个行动的效果以及整体框架观, 254-5

hypothetical motivational sense of 'could' (“能够”的假设性动机意义): 260, 这个意义与“应当蕴含能够”是相关的, 261-3

ideal deliberation, fully informed and procedurally rational (理想的慎思, 信息完全和程序上理性的): 62-4, 我们能从理想慎思的主张中获得什么, 78-80, 83-4, 93-5, 103-4

Ideal world objections (理想世界反驳): 18; the old objection (古老的反驳), 312-15; the new objection (新反驳), 315-17; one solution (一种解决方案), 317-19; 它提出我们在此会忽略的复杂性, 341, 479 (on Ridge)

Imagined cases (虚构的各种情形):

Aid Agency (援助机构情形), 同意原则是否要求过高: 209-10

Bad Samaritan (不乐善好施者), 把人们不是纯粹当作手段, 而是纯粹当作物件来对待: 226-7

Blind man crossing the street, and ought implies can(盲人过街和应当蕴含能够): 260-1, 108

Blue's Choice, and Rawls's thin theory of the good (布鲁的选择和罗尔斯有关善的薄理论): 104-6

Bomb, and the priority of our duty not to kill (炸弹情形和我们不杀人义务的优先性): 449; 与隧道情形的对比, 450

Bridge (桥梁情形): 作为手段的伤害和纯粹作为手段的对待: 218-220, 223-4, 228-30; and Scanlonian Contractualism(桥梁情形和斯坎伦式契约论), 364-6; and the Wrong-Making Features Objections to the Kantian Argument (桥梁情形以及对康德式论证的致不当特征反驳), 393-7; and deontic reasons (桥梁情形和道义式理由), 450-1

Depression, and justified coercion (抑郁情形和有证成的强制) 197-8

Desperate Plight (绝望困境情形), 把某人纯粹当作手段来对待: 225

Drowning Child (溺水的孩子): 利己主义者救了他并将其纯粹当作工具, 但其行动并无不当, 216, 正如第三种纯粹手段原则所承认的, 217; 康德的普遍法则公式谴责这种行动为不当的, 290-1

Earthquakes (地震情形): First Earthquake (第一种地震情形), 作为对同意原则的检验, 185-6, 192-3, 201-2, 208; Second or Means, (第二种地震情形或手段情形) 用来阐明给予同意的道义式理由, 201-4; Third and Fourth (第三种和第四种地震情形) 作为纯粹手段原则的检验, 222-3, 231-2

Equal Claims (平等的要求权), 我们是否能理性地同意只给予得救的均等机率: 206-7

False Belief (虚假信念), 耶和华见证会教徒能够有效地拒绝同意: 198

Fatal Belief, and justified deceit (致命的信念和有证成的欺骗): 178

Fire, and deontic reasons to give consent (火灾情形与给予同意的道义式理由): 204-5

Future Tuesday Indifference (对未来周二的漠不关心): 56, 79, 120-4

Gangster, and treating people merely as a means (土匪情形, 以及把人们纯粹当作手段来对待): 216, 228, 231-2

High Price and the Consent Principle (高价格情形和同意原则): 207

Lesser Evil (较小的恶情形), 某人行动不当如何可以是更好的: 393, 475; 较小的恶以及对于康德式论证的一种反驳, 393-4, 399

Lifeboats (救生艇情形): 第一种救生艇情形作为同意原则的检验, 186-8; 作为康

德式契约论公式的检验, 380-4; 第二、三和四种救生艇情形作为这个公式的进一步检验, 385-9

Means (手段情形), 以及给予同意的道义式理由: 201-4, 209

Mine Shafts (矿井情形), 以及做期望上最好的事情: 159-60

Mistake, and the Ideal World Objection (失误情形, 以及理想世界反驳), 313-4; 遵循行动后果论原则是否总是能使事情进展最好, 405-6, 482-3

Murderous Theft, and the Rarity and High Stakes Objections to the Kant's Formulas (谋杀性偷窃情形, 以及对康德的公式的稀罕性反驳与大赌注反驳): 331-3

Mutual Benefit (互利的情形), 把某人纯粹当作手段来对待: 217

Parents, and the Consent Principle (父母情形, 与同意原则): 205

Rescue (救援情形), 我们如何做最好的事情: 253, 256

Scarlet, Crimson, and Pink irrationality and inconsistency (斯加力, 克里蒙森与平克, 不理性与不一致): 120-4

Schelling's Case, and rational irrationality (先令的情形, 以及合理的不合理性): 437

Self (自我情形), 以及同意原则是否要求过高: 207, 211

Self-Defeating Desire (自我挫败的欲望), 以及所谓的状态给予的理由: 431

Self-Defence (自卫的情形), 以伤害某人作为手段但不用把这个纯粹当作手段来对待: 221-2, 225

Shipwrecks (船舶事故情形), 救自己还是救陌生人: 139-41

Surgery (手术情形), 有证成的强制: 196-7

Transplant, Act Utilitarianism and Scanlonian Contractualism (移植情形, 行动效用主义和斯坎伦式契约论): 363-5, 473 (on 363)

Tunnel (隧道情形): 218; 我们不杀人的义务优先于我们救人的义务, 对此信念的一个反例, 450; 与桥梁情形的比较, 219-220, 228-30, 364-6, 449-51

Unjust Punishment, and the Rarity Objection to Kant's Formulas (不公正的惩罚情形, 以及对康德的公式的稀罕性反驳): 330-1

Whimsical Despot (反复无常的专制者), 对理由的非自愿回应 47; 以及欲望的合理性, 125-6; 以及状态给予的理由, 420-30; 以及主观主义, 457; 以及对康德式契约论公式的一种反驳, 476 (Rosen's malicious gremlin [罗森之恶意的精灵])

impartial reasons (不偏不倚的理由): 参见 Reasons (理由)

Impartial-Reason Act Consequentialism as a rival to morality (不偏不倚理由的行动后果论是取代道德的): 168-9, 171, 500 (on 169)

Impartial Observer Formula (不偏不倚的观察者公式): 可能诉诸不偏不倚的观察者之合理选择, 329-30, 或者诉诸其事实上的选择, 458 (on 101)

impartiality, ways of achieving (实现不偏不倚的各种方式): 具有一种不偏不倚的视

角, 40-1, 133-4, 这种视角不是必要的, 135-7; 遵循黄金规则, 321-30; 设想过他人的生活, 325-6; 遵循同意原则, 329; 康德的普遍法则公式, 333-8; 康德式契约论公式, 339-41; 理性协议公式, 343-6; 无知之幕和罗尔斯式契约论 348-51, 356-7, 472-482, 和康德式契约论: 383-4, 389; 401-2

imperatives, hypothetical and categorical (假言命令与绝对命令): 243

impersonally good, as contrasted with good for (与对……是好的相对比的非个人地好的): 41-2

Importance (重要性): 心理重要性, 97; 在规范的、蕴含理由意义上的重要性, 146-8; 以及参见 *mattering* (重要)

Impossibility Formula (不可能性公式), 参见 *Kant's Formulas* (康德的各种公式)

imprecise comparability (不精确的可比性): 33, 132, 137-9; 以及卷二

Incoherence Argument against Subjectivism (反主观主义的不融贯论证): 91-101

incompatibilism about the freedom required for morality (有关道德要求的那种自由的不相容论), 258-263; about the freedom required for deserving to suffer (有关应得痛苦所要求的那种自由的不相容论), 263-272

Inconsistency (不一致性): 欲望的不一致, 126-8; 信念的不一致, 128-9; 不一致性与理性要求, 36, 118, 123

Indirect Consequentialism (间接后果论): 22, 374-5

individual rationality (个人理性), 在群己困境中并未表明是自我挫败的, 306

Informed Desire Theory of reasons (理性的知情欲望论): 61, 94-6

instrumental desires (工具性欲望), 44; 工具性欲望不能言之成理地给予我们理由, 59; instrumental reasons (工具性理由), 52, 从内在理由获得其效力, 90-1

instrumental rationality (工具合理性): 125, 135, 243, 497 (on 510)

intensity of pain (痛苦的强度), 心理学意义与规范意义: 132, 459-60

intentional objects of desires, and the rationality of these desires (欲望的意向对象以及这些欲望的合理性): 112, 129

intentionally doing, morally relevant descriptions of an act (意图做的, 行动在道德上的相关描述): 297-8, 466

interactionist dualists (互动论二元论者): 265

internal senses of 'reason, 'should, and 'ought'(理由、“应该”和“应当”的内在意义), 72; 以及卷二

Internalism about Reasons(理由的内在论): 参见 *Reasons*(理由), *Subjectivism*(主观主义); 以及卷二

intransitive and transitive relations (传递与非传递性关系): 128, 459

intrinsic goodness (内在善性): 38-9, 42; 和工具善性, 50; 外在善性, 52, 455; 与许多主观主义者的信念, 93-101; 与欲望的合理性, 129, 497 (on 107); 事件与事物的善性, 236-7

intuitive beliefs or intuitions (直觉信念或直觉): 366–8, 370, 185, 346, 362; 以及卷二

Irrationality (不合理性): 参见 Reasons and Rationality (理由和合理性)

irreducibly normative truths (不可还原的规范真理): 109–10, 494 (on xlv); 以及卷二

irreversible consent (不可逆的同意): 193–6, 202

rightness as fairness (作为公平的正当性): 罗尔斯的契约论道德理论, 346 and 511

Justice (正义): 参见 distributive justice (分配正义) 和 desert (应得)

justifiabilist senses of ‘ought’ and ‘wrong’ (“应当”与“不当”的证成意义): 166, 170, 174, 368–9

Kagan, Shelly (谢利·卡根): xlvi

Kahane, Guy (盖伊·卡恩): xlvi

Kamm, Frances (弗朗西斯·卡姆): 论把人们纯粹作为手段来对待, 213, 501–2; 论作为手段的伤害 464 (on 229), 492 (on 450); 论最优的义务论禁令, 478; 论给予人们均等的机率 463; 论距离的道德相关性, 514

Kant (康德): 与西季威克的比较, xxxiii–xxxiv; 他的伟大, 以及我们为什么应该读其著作, xli–xliv; 有关康德主张的提要, 8–26

Kant’s Formulas (康德的各种公式):

Humanity Formula (人性公式): 177, (and 500); 康德有关同意的主张, 177–81; 同意原则与康德的理想): 8–10, 180–211, 483–4; 把人们不纯粹作为手段, 也作为目的来对待: 10–12, 212–28, 463; 作为手段的伤害, 228–232; 尊重人, 233–5; 康德有关善良意志、理性存在者、理性、目的王国和最伟大善之理想世界(作为自在目的或有待产生的目的)所具有的价值、尊严或最高价值之主张, 12–13, 235–44, 464

Greatest Good Formula (最伟大善公式): 康德的后果论, 13, 244–57; 康德是否犯有他称之为古希腊人的“根本错误”, 243–49; 我们应当如何总是致力于促进最伟大的善 249–57; 通过遵循康德的其他公式, 250–1; 康德有关幸福的主张, 243; 享乐主义的规则效用主义, 255–6, 409–10 (and 484); 与应得、道德与自由意志, 13–14, 257–272

Universal Law Formula (普遍法则公式): “我应当永远只按照也能意愿其变为普遍法则的那种准则行动”(G 402): 不可能性公式, 14, 275–84, 465–6

Permissibility Formula (可允许性公式), 286, 466, 508

Law of Nature and Moral Belief Formulas (自然法则以及道德信念公式): 15, 284–88, 508, 466–7; 意愿作为普遍法则, 285–8, 301–2; 我们应用这些公式时应

该做的假设：我们不应该诉诸有关何种行动不当的信念，287-8，也不应该诉诸有关合理性的某种信念，288；相关的替代项，301-2, 466；当自然法则公式最有成效时，17, 301-8, 467-9

Mixed Maxims Objection (混合型准则反驳)：行动之不当性不能取决于行动者的准则，15-16，我们应如何描述行动者的准则，289-90；守诺、还债且救溺水儿童之利己者的准则，290-2；康德“永不撒谎”之准则，292；康德的公式应如何修正，16-17, 293-300, 466 (on 298)

Threshold Objection (门槛反驳)，18, 308-312；理想世界反驳，18, 312-20, 341, 469, 479

Maxim Consequentialism and Kant's Law of Nature Formula, the Relativism Objection (准则后果论和康德的自然法则公式，以及相对主义反驳)，375-7

Kant's Formulas and the Golden Rule (康德的各种公式和黄金规则)，321-30；稀罕性与高筹码反驳：20, 289-90, 296-7, 330-33

Non-Reversibility Objection (不可逆反驳)：19-20, 334-8；康德的公式再次应如何修正，变成康德式契约论公式：20-1, 338-42, 470-1；以及参见 Contractualism, Kantian (康德式契约论)，也参见卷二

Kantian Argument for Rule Consequentialism (对规则后果论的康德主义论证)：参见 Contractualism, Kantian (康德式契约论)

Kantian view of blameworthiness (可责备性的康德式观点)：155-7, 461-2

Kemp Smith, Norman (诺尔曼·坎普史·密斯)：xlii, 464, 494

Kolodny, Niko (尼可·柯洛笛)：xlvi；论理由和合理要求，495；论涉及爱的理由，496 on (73)

Korsgaard, Christine (克里斯汀·科丝嘉)：xlv；论善性的主观主义和客观主义，46, 55, 94；论基于欲望的理由，65；论康德有关同意的主张，177-80, 501；论康德有关把他人纯粹作为手段来对待的主张 224-6, 502；论康德的不可能性公式，278, 282；提出此公式的另一版本，284；论对康德公式的理想世界反驳，313, 510；论康德的普遍法则公式，337；论康德有关人性的主张，464，以及论自杀，xliii, 494

Law of Nature Formula (自然法则公式)：参见 Kant's Formulas (康德的各种公式)

Lenman, James (詹姆斯·雷曼)：481-2

Lesser Evil (较小的恶情形)，某人行动不当如何可以是更好的：393, 475；较小的恶以及对于康德式论证的一种反驳，393-4, 399

Lifeboats (救生艇情形)：第一种救生艇情形作为同意原则的检验，186-8；作为康德式契约论公式的检验，380-4；第二、三和四种救生艇情形作为这个公式的进一步检验，385-9

Love (爱)：作为福祉的一部分，39；欲望与涉及爱的理由，73, 496 (on 73), 141；爱作为理由的一种来源，98；我们爱人的理由，100-1；爱我们的敌人，126；爱与那些

与我们具有紧密纽带的人, 136; 爱与具有充分的理由行动不当, 143; 爱某人但将其纯粹作为手段, 215; 内在地善的行动, 236; 爱与我们的幸福, 251, xxxvii; 爱与对规则后果论的康德式论证, 387-9; and 爱与这样一种方式, 如果我们大部分人要么接受要么遵循行动后果论原则, 事情会变得更糟, 404, 406, 387 (‘深刻的情感……’)

lying promises (许假诺): 许假诺与同意的不可能性, 177; 许假诺与把他人纯粹当作手段来对待, 216; 许假诺与康德的不可能性公式, 279-81

Marginalist View about how we can do the most good (我们如何做最好事情的边际主义观点): 253-6

masochism and the Golden Rule (受虐狂与黄金规则): 324

Mattering (重要): 心理意义与规范意义, 96-101; 涉及理由的意义, 144-8; 论主观理论, 尽管事情对人们是重要的, 但没有任何事情重要, 106-7, 110; 论有关何物作为道德的外在对手而重要的信念, 169; 道德是否以及有多重要, 172-4; 重要与后果论, 417; 现在最重要的事物, 419; 以及卷二

Maximin Argument (最大的最小值论证): 352-4, 472

maximizing happiness (幸福的最大化), 去掉痛苦后产生的幸福总量最大: 169, 251, 463

Maxims (准则): 康德的意思以及他的一些例子, 275; 准则应如何描述, 289-90; the Mixed Maxims Objection (混合型准则反驳): 行动的不当性不可能取决于行动者的准则, 15-16, 289-93; 康德的公式谴责的、其他好的或可允许的准则, 308-12; 这些公式没能谴责的坏准则, 315-20, 335; 准则与康德式契约论公式, 471 (on 342); 准则后果论与康德原本的普遍法则公式, 375-7; 行动后果论准则是否可能被意愿为普遍的, 482-3

Means (手段), 以及给出同意的道义式理由: 201-4, 209

means, treating merely as (纯粹当作手段来对待): 10-12, 212-28, 463-4; 不同于作为手段的伤害, 228-232

Mere Means Principles (纯粹手段原则): 第一种与第二种纯粹手段原则, 212-214; 第三种纯粹手段原则, 217-21, 228-232

meta-ethical and meta-normative theories (元伦理理论和元规范理论): 109-10, 174, 367; 以及卷二

meta-hedonic desires 元享乐式欲望, 很是不同于享乐式喜恶: 54-65

Metaphysical Naturalism (形而上学自然主义): 109-10; 以及卷二

Mine Shafts (矿井情形), 以及做期望上最好的事情: 159-60

Mistake, and the Ideal World Objection (失误情形, 以及理想世界反驳), 313-4; 遵循行动后果论原则是否总是能使事情进展最好, 405-6, 482-3

Mixed Maxims Objection to Kant's Formula of Universal Law (对康德普遍法则公式的混合型准则反驳): 15-16, 289-93

Moore, G. E. (GE 摩尔): (善性与存在), 237; “正当”意指“使事情进展最好”, 247, 464–5; 论遵循最优规则, 252; 论性欲的快感, 453

Moral Belief Formula (道德信念公式), 参见 Kant's Formulas (康德的各种公式)

moral-belief-relative senses of 'ought' and 'wrong' “应当”和“不当”的道德信念相对的意义: 158–161

moral luck, and agent-regret (道德运气以及行动者遗憾): 156–7, 461–2

Moral Rationalism (道德理性主义): 141–4

moral status, value, dignity, and worth, Kant's claims about (康德有关道德身份、价值、尊严的主张): 235–44

moral theories, different parts of (道德理论的不同部分): 407

moral worth (道德价值): 道德价值与行动的不当性, 217, 282, 290–1; 道德价值与康德的普遍法则公式, 293, 299–300; 道德价值与善良意志, 240–50

moralist's problem (道德主义者的问题): 143

morally-decisive-reason senses of 'ought' and 'wrong' (“应当”和“不当”之道德上的决定性理由的意义): 167, 170

morally relevant facts, or descriptions of acts (道德上相关的事实或行动的描述): 294–5, 298, 466 (on 298)

Morgan, Seiriol (西尼尔·摩根): 479–81

morally responsible, in the suffering deserving sense (在应得痛苦的意义在道德上负责), 参见 desert (应得)

motivating reasons (驱动性理由): 37; 驱动性理由可用两种方式描述, 454–5; 驱动性理由和主观主义, 66, 107–10; 驱动性理由和自由意志, 266

Motive Consequentialism (动机后果论): 动机后果论和系统的德性伦理学, 375; 个体主义版本, 475; 作为更广泛理论的一部分, 407

M-related people, and common sense morality (M 相关者与常识道德): 304–7, 468–9

Murderous Theft, and the Rarity and High Stakes Objections to the Kant's Formulas (谋杀性偷窃情形, 以及对康德的公式的稀罕性反驳与大赌注反驳): 331–3

mustn't-be-done, the indefinable sense of 'wrong' (绝不可做, “不当”的不可定义的意义): 165–6, 169–70, 173, 451

Mutual Benefit (互利的情形), 把某人纯粹当作手段来对待: 217

Nagel, Thomas (托马斯·内格尔): v, xlv; 论个人视角和不偏不倚的视角, 133, 498–9, 136–9; 论黄金规则, 325–6; 论康德的普遍法则公式, 338–9; 论行动者遗憾, 461; 捍卫可责备性的准则康德式观点, 500; 罗尔斯的评价, 471 (on 346); 不可还原的规

- 范真理, 494 (on xlv); 论元伦理学, 174; 以及卷二
- Naturalism, Metaphysical (形而上学的自然主义): 109–10; 以及卷二
- Naturalistic Fallacy (自然主义谬误): 既不是自然主义的也不是谬误, 西季威克的理
解比摩尔好, 465
- Nearness Principle (就近原则): 381–2, 514
- New Ideal World Objection (新理想世界反驳): 316, 341
- Newman, Cardinal (卡笛莱尔·纽曼) 论痛苦与罪的相对坏性质: 241
- Nietzsche, Friedrich (弗里德里希尼采): 54, 以及卷二
- No Knowledge Formula, and Rawlsian Contractualism (无知公式和罗尔斯式契约
论): 350–2
- No-Agreement World, and Hobbesian Contractualism (无协议世界与霍布斯式契约
论): 344–5
- 非道义式理由: 拒绝同意, 201; 与康德对规则后果论的康德式论证, 390–98, 448–
51
- non-moral goodness and badness (非道德的好与坏): 38–42, 243, 371–2
- Non-Reversibility Objection (不可逆反驳): 19, 334–8, 341
- non-voluntary responses to reasons (对理由的非自愿回应): 47–50, 117–18, 420–
32
- normative disagreements (规范性分歧): 418, 以及卷二
- Normativity 规范性: 规范性概念: 非道德的, 31–42; moral, 150–74; 涉及理由和涉
及规则的观念, 144–8, 驮负 mattering (重要); 规范性和动机, 107–10; 实质性的
规范主张, 70, 参见 (concealed tautology) 隐藏的同义反复; 规范性力量, 35; 规范
性与派生理由, 39, 66, 172; 因果地依赖但规范上不依赖欲望的理由, 68; 工具性理
由, 90–1; 不可还原的规范真理与形而上学自然主义, 109–10, 以及卷二
- noumenal beings (本体式存在): 258, 242, 263, 269, 351
- Nozick, Robert (罗伯特·诺齐克): 基于价值的理论和“专制要求”, 66; 忽略对象
给予的理由, 497; 论纯粹作为手段来对待, 502 (on 217)
- Numbers Principle (数字原则): 380–3, 388–9, 397
- O'Neill, Onora (奥诺拉·奥尼尔): 康德“最令人气愤的”一本著作, xlii; 论欺骗、
强制和同意的可能性, 177–80; 论纯粹作为手段来对待, 224–6; 观念中的矛盾检验
465 (on 275); 提出康德不可能性公式更弱的版本, 283–4; 论我们应如何描述某个
行动者的准则, 289–90; 论康德的普遍法则公式背后的“直观思想”, 328; 表示这
个公式是想告诉我们只有何种行动才具有道德价值, 293; 但是康德的公式无法实现
这个目标, 299–300
- oaths, and arguing from 'is' to 'ought' (咒骂与从“是”到“应当”的论证): 280
- Objectivism about reasons (理由的客观主义): 参见 Reasons (理由)

Optimific (最优的): 使事情进展最好: 不同的意义, 375, 377, 405, 475; 参见
Consequentialism (后果论)

Otsuka, Michael (迈克尔·奥特苏卡): xlvi, 478–9

ought implies can (应当蕴含能够): 107, 258–9, 438

ought and should (应当与应该): 在蕴含决定性理由的意义上 33; 不同意义的实践
应当: 162–3; 理性上应当, 33–6, 163–4; 应当不偏不倚地, 168–169; 参见 Reasons
(理由)

ought morally (道德上应当): 参见 wrong (不当的)

ought epistemically (认知上应当): 117–8, 426

ought-based sense of 'good' (“善”基于应当的意义): 247–9

Pacifism (和平主义): 312–15

pain, agony, suffering (疼痛、痛苦和苦难): 2–4; 相关意义的“疼痛”, 53–54,
455–6; 一种拓展广义“疼痛的”, 226; 其坏处和避免未来疼痛的理由, 31, 56–7,
129, 138; 为基于欲望的主观理论所否认, 73–7, 81–9, 456–7; 也许为斯多葛学派
所否认, 371, 513; 为康德与罗斯所忽略, 371–2, 513; 与罪的坏性质相比, 241; 快
乐与痛苦的强度与持续时间, 其相对的重要性, 132, 强度的心理学意义与规范意义
459–60; 痛苦与偏向未来, 197; 福祉的享乐主义理论 39, 105; 动机中的, 44; 论合
理性或道德, 169, 246; 是否能应得痛苦 257, 264–72, 409–10; 以及卷二

Parents, and the Consent Principle (父母情形, 与同意原则): 205

完善论: 389

Permissibility Formula (可允许性公式): 参见 Kant's Formulas (康德的公式)

person-relative and partial reasons(个人相对和偏向性理由): 40, 以及参见 Reasons(理
由)。

Persson, Ingmar (英马尔·佩尔森): xlvi

phenomenal world (现象世界): 258, 269

Pleasure (快乐): 以及欲望, 心理享乐主义, 44–5; 快乐和享乐式喜恶, 53–6; 快乐与
有关价值与理由的主观主义, 55, 67–8; 性快乐, 西季威克和摩尔, xxxviii, 453; 以
及参见 happiness, pain, and Reasons (幸福、痛苦和理由)

Pogge, Thomas (托马斯·博格): 以及对康德公式的门槛反驳, 309–10

practical reasons (实践理由), 参见 Reasons (理由)

Principle of Equal Shares (均等份额原则): 345, 359–60

principle of self-love (自爱原则): 291–2, 513 (on 371)

procedural rationality (程序合理性): 62–3, 78–80, 103, 347, 496

profoundest problem (最深刻的问题): 西季威克的: 6–7, 130–49, 498 (on 133),
499 (on 143), 461 (on 143); 这个问题的宽泛形式, 144–8

progress, philosophical and moral (哲学和道德的进步): xxxiii, 174; 以及卷二

Promises (诺言): 以及康德的不可能性公式, 279–281, 465; 对亡者的诺言, 374; 常识道德与联合的有条件的诺言作为解决某些道德上的群己困境的方案 468; 高契尔的理论, 433–445; 485, 488–91, 应得与幸福的相称性条件: 245

Psychological Egoism 心 (心理利己主义): 43

Psychological Hedonism (心理享乐主义): 44

punishment, justification of (惩罚的证成): 455, 461–2; 以及卷二

Rarity Objection (稀罕性反驳): 289–90, 296, 330–31

Rational Agreement Formula, and Hobbesian Contractualism (理性协议公式和霍布斯契约论): 343–6, 348, 355–7

Rational Egoism (理性利己主义): 52; 有些版本是隐藏的同义反复, 125, 135; 实质的理性利己主义不是一种道德观点, 而是道德的外部对手, 166; 利己主义的卑鄙性, 169; 但不可否认是说得通的, 498 (on 133), 499 (on 143); 利己主义和非利己主义理由, 130–44, 358–9; 利己主义和群己困境, 306–7; 利己主义和慎思主观理论, 347–8; 作为对康德式论证的反驳的高筹码利己主义, 383–4; 利己主义和高契尔的理论, 433–47

Rational Impartialism (理性的不偏不倚主义): 52, 130–1, 168–9; 参见 Consequentialism (后果论) 和 rationalist's problem (理性主义者的问题): 143

Rationality (合理性): 参见 Reasons and Rationality (理由与合理性)

Rawls, John (约翰罗尔斯): 有关合理性与理由的信念: 78, 144; 稀薄的善理论, 103–5; 论康德有关同意的主张, 182–3; 论正当与善, 235; 论如何应用康德的普遍法则公式, 279, 288; 论不自杀作为一种德性之责, 290; 提出康德假定了一种无知之幕, 339; 论黄金规则, 329; on 论重新定义“正确”与“真”, 226–7; 342; 论道德理论, 174; 作为公平的正当性, 罗尔斯的契约论道德理论, 346, 511: 参见 Contractualism, Rawlsian (罗尔斯式契约论)

Raz, Joseph (约瑟夫·拉兹) (其观点是我本该讨论的): 495 (on 65)

reactive-attitude sense of 'wrong' (“不当”之反应性态度意义): 165, 169–71, 174; 道德指责不同于希望某个人过得不好, 或不再希望其过得好, 272

Reasons and Rationality (理由与合理性): 1–8, 27–8; 31–149

the concept of a reason, (一个理由这种概念) 31; 有一个理由和是一个理由, 32

practical reasons, (实践理由), 广义上的“实践的”, 45, 47, 65

sufficient and decisive reasons, most reason (充分的理由、决定性理由和最有理由): 32–3, 454

给予理由的事实: 34–7, 42, 111; 有些人称它们为: 是理由的事实 32; 主观主义者不能诉诸, 94–5

the reason-involving concepts should, ought, must (涉及理由概念的应该、应当和必须 33), 454 (on 33); 好与(坏) 38–9; 对……是好的, 和非个人地好的,

41-2; “非个人的”的这种用法可引起误解, 41-2

apparent reasons (显见理由), 参见下面的 Rationality (合理性)

motivating reasons (驱动性理由): 37; 可用两种方式描述; 454-5

telic, instrumental, intrinsic, and extrinsic reasons (目的式、工具性、内在的和外在的理由): 52, 455

Objectivism about reasons (理由的客观主义): 2-3, 5-7

object-given reasons (对象给予的理由), 45-7

the sense in which these reasons are value-based (理由是基于价值的那种意义), 455

reason-involving kinds of goodness (涉及理由的各类善性), 38-42, 101-2

hedonic reasons are not desire-based but object-given (享乐式理由不是基于欲望而是对象给予的), 52-6, 81-2, 456-7

different objective theories (不同的客观理论), 45-7, 52, 130

partial and personal reasons (偏向性的和个人性的理由), 40, 135-138, 143, 379, 389; 以及参见 Rational Egoism (理性利己主义)

impartial reasons (不偏不倚的理由): 6-7, 22-3; 40-2; 蕴含不偏不倚理由意义的“不偏不倚地最好的, 41-2, 主观主义者不能使用这种理由, 101-3; 我们也许有不偏不倚的理由更关注某些人的福祉, 41; “应当”和“不当”之蕴含不偏不倚理由的意义, 167-71, 372; 给予同意的不偏不倚理由, 187-8; 选择某个普遍原则的不偏不倚的理由 378-88, 391-403; 不偏不倚的理由和个人相对的道义式理由 246, 475 (on 393)

conflicts between partial or personal and impartial reasons (偏向性或个人性理由与不偏不倚理由之间的冲突): 130-44; Sidgwick's Dualism of Practical Reason (西季威克的实践理性二元论), 131-4, 499 (on 143); 两视角论证, 134-7; 广义的基于价值的客观论, 136-141

reason-involving conception of normativity (涉及理由的规范性观念), 144-5; 事物重要的那种涉及理由的意义, 146; 理由比“理性或道德要求”更为根本, 145-8

reasons and morality (理由与道德): 141-4, 147-9; 道德理性主义, 141; 西季威克的自我利益与义务的二元论, 142-4, 460 (on 142); 弱道德理性主义, 144; 修正后的最深刻问题, 144-5, 147-8

deontic reasons (道义式理由): 道义式理由与同意原则, 201-2; 道义式理由与康德的公式, 287-8; 道义式理由与非道义式理由, 390, 395, 448-51

derivative reasons (派生理由): 39; 派生理由与善性, 42

epistemic reasons (认知理由): 持有信念的真值相关的理由, 47-51; 认知理由与实践理由如何可以竞争但不可能相冲突, 425-8; 认知理由与形而上学的自然主义: 110 以及卷二

- 我们是否能有状态给予的或实践的理由持有欲望或信念：50-1, 420-32, 442-4
- Subjectivism about reasons (理由的主观主义)：1-5; 45-7, 58-109
- desire-based, aim-based, and deliberative theories (基于欲望和目标的慎思理论), 58-65, 456
 - unlike desire-based theories of well-being, subjective theories about reasons appeal only to facts about present desires (不同于基于欲望的福祉理论, 理由的主观理论只诉诸于有关当下欲望的事实), 74-5
 - how the Deliberative Theory may seem Objectivist; procedural and substantive rationality (慎思理论如何可以看来是客观主义的; 程序合理性与实质合理性), 62-3, 79-80
 - why so many people accept subjective theories (为什么这么多人接受主观理论), 65-70
 - Analytical and Substantive Subjectivism (分析性主观主义与实质性主观主义), 70-3
 - Subjectivist claims about self interested and moral reasons (有关自利理由和道德理由的主观主义主张), 102; on subjective theories, we may have no reasons to do our duty, and decisive reasons to act wrongly (基于主观理论, 我们可能没理由履行我们的义务, 有决定性的理由行动不当) 144, 347
- Arguments against Subjectivism (反对主观主义的论证)：痛苦论证, 73-82, 456-7; 主观主义者不能以不考虑虚构情形来回应这个论证, 76-7; 也不能诉诸有关程序合理性的主张来回应这个论证, 77-81; 全部或全不论证：83-91; 尽管很难设想为其自身的缘故而想要痛苦, 这并没削弱这个论证, 83-4; 我们可能没有任何基于欲望的理由不满足这种欲望, 85-9; 主观主义者不可能只诉诸我们我们有理由持有的欲望, 89-91, 106-7
- the Incoherence Argument (不融贯论证)：91-6; 那些诉诸知情欲望的人并不是真正的主观主义者, 91-6; 法兰克福的观点, 96-101
- Subjectivists cannot make positive normative claims about the goodness of outcomes (主观主义者不能对结果的善性提出正面的规范主张), 101, 或者是对福祉提出正面的规范主张, 101-6, 496-7, 或者是对何物重要提出正面的规范主张, 107
- Arguments for Subjectivism (主观主义的论证)：
- 诉诸享乐式理由, 67, 53-6,
 - 诉诸派生理理由, 66 ('Fifth...')
 - 诉诸“应当蕴含能够”和驱动性理由, 108-10
 - 诉诸分析性主观主义：72-3;
 - 诉诸规范性的动机论说和形而上学自然主义：70-3, 以及卷二
- Rationality (合理性)：
- senses of 'rational' and 'irrational' (“合理”与“不合理”的意义)：日常意义,

33; 斯坎伦的狭义, 123; 基于当前欲望的意义和利己主义意义, 123, 135

when we are aware of facts that give us decisive reasons, we ought rationally respond to these reasons (当我们意识到事实给予我们决定性的理由时, 我们应当理性地对这些理由做出回应), 111; 如果我们没有意识到给予理由的事实, 我们就不会在回应上失败, 32, 111; 不同于对行动理由的回应, 我们对于持有欲望或认知理由的回应很少是自愿的, 47–51, 420–24

rational responses to conflicting reasons (对冲突理由的合理回应), 130–6; 广义的基于价值的观点 138–41, 186–7, 382–3, 460; 合理性与道德, 141–9

some examples of irrational desires or preferences (几个不合理欲望或偏好的例子), 55–7, 79, 83–4, 104; 持有不合理信念和欲望的理由, 合理的不合理性, 125–6, 420–32, 439–441

though reasons are given only by facts 尽管理由是只由事实给予的 (但是参见 454, on 35), 但是我们应当理性地想要或做的事情取决于我们的信念, 33–5; 我们为什么应做出这个区分, 36; 当我们持有的信念若为真就会给予我们某种理由时, 我们具有显见理由; 显见理由可以是真实的理由也可以只是显见的: 34–5, 111–13; 当我们持有某些欲望或做某些事情是因为我们具有不合理的非规范信念时, 我们的欲望或行动是否应被称为不合理的, 113–17

what we ought practically to do in the fact-relative, evidence-relative, belief-relative, and normative-belief relative senses (在事实相对、证据相对、信念相对和规范信念相对的意义上我们实践上应当做什么): 162–3; 有关风险和不确定性的问题, 37, 125, 159–63; 期望主义, 160, 462–3

epistemic rationality (认知合理性): 110–120; 在认知合理性与实践合理性之间以另一种方式做出的更深刻的区分 116–18, 427

rational requirements and inconsistency between our normative beliefs and other mental states (合理要求与“我们的规范信念和其他心理状态”之间的一致性, 36, 118–25; 其他的合理要求, 36, 146; 我们遵循这些要求的理由, 146–7, 495 (on 36); 有关合理性的其他观点, 125–9, 135; 工具合理性: 90–1, 125, 497 (on 107); 程序合理性与实质合理性, 62–3, 79–80; 慎思理论, 62–3, 103–6, 126

the existence of rational beings (理性存在者的存在), 31, 419

reasonable rejection, in Scanlonian Contractualism (斯坎伦式契约论中的合情理的拒绝): 360–5, 416, 473

Reid, Thomas, (托马斯·里德), “成为无赖或傻瓜是否是最好的: 142

Relativism Objection (相对主义反驳): 377

religion: and Sidgwick’s profoundest problem (宗教: 和西季威克最深刻的问题), 142; 宗教与有效同意, 198; 宗教与相信道德不是一种幻觉, 259; 宗教与道德分歧, 418

Rescue (救援情形), 我们如何做最好的事情: 253, 256

respect for persons (对人的尊重) : 233–5

responsible, in the suffering-deserving sense(在应得痛苦的意义上负责) : 参见 desert(应得)

retributive justice (报应性正义) : 参见 desert (应得)

retroactive endorsement (回溯性认可) : 196–8

Rich (富裕者) :

 we rich people (我们富裕者) : 我们最重要的道德问题, 501 (note about 210); 康德的公式, 337; 富裕者与同意原则, 209–10; 富裕者与规则后果论, 319; 富裕者与重要的事物, 419

Ridge, Michael (迈克尔·里奇) : 469 (on 319) and 479

rightness as fairness, Rawls' contractualist moral theory (作为公平的正当性, 罗尔斯的契约论道德理论) : 346, and 511 (on 346); 以及参见 Contractualism, Rawlsian (罗尔斯式契约论)

Rights Principle (权利原则) : 194–7

risk and uncertainty(风险与不确定性) : “应当”与“不当”的重要性以及不同的意义, 37, 125, 159–63; 期望主义, 160, 462–3; 与罗尔斯对无知之幕的诉诸, 349–53, 357, 472–3

Rosen, Gideon (吉迪恩·罗森) : 476

Ross, Sir David (大卫·罗斯爵士) : 372, 464 (on 241)

Ross, Jacob (雅可布罗斯) : xlvi, 476–8

Rule Consequentialism (规则后果论) : 375, 以及参见 Consequentialism (后果论)

rule-involving conception of normativity (涉及规则的规范性观念) : 144–5

Ruskin, John (约翰·罗斯金) : 与康德的比较 : 453

Samaritan's Dilemmas (乐善好施者的困境) : 467

Sartre, Jean-Paul (让-保罗·萨特) : 使用的一个例子并不支持其观点, 458 (on 100)

Scanlon, T. M. (T. M. 斯坎伦) : v, xlv, 其有关涉及及理由之善性的主张, 39, 495; 论自主性和其他人的欲望, 66, 496; 论理由与合理要求, 495 (on 146); 斯坎伦对康德的普遍法则公式的诠释, 339–41, 以及对可允许性公式的诠释, 508 (on 286); 论合理性和有关理由的信念, 119–125; 提出我们在狭义上使用的“不合理的”, 495 (on 65); 论要被尊重而不是促进的价值, 236–9; 斯坎伦的道德理论, 参见 Contractualism, Scanlonian (斯坎伦式契约论)

Scarlet, Crimson, and Pink irrationality and inconsistency(斯加力, 克里蒙森与平克, 不理性与不一致) : 120–4

Schelling's Case, and rational irrationality (先令的情形, 以及合理的不合理性) : 437

Schneewind, Jerome (杰罗米·席温德) : 493

Self (自我情形), 以及同意原则是否要求过高 : 207, 211

self-creation, the impossibility of (自我创造的不可能性): 267–9

Self-Defeating Desire (自我挫败的欲望), 以及所谓的状态给予的理由: 431

Self-Defence (自卫的情形), 以伤害某人作为手段但不用把这个纯粹当作手段来对待: 221–2, 225

self-love, the maxim or principle of (自爱的准则或原则): 275, 291–2, 513

Semi-Kantian view of blameworthiness (可责备性的准康德式观点): 56

separateness of persons and distributive justice (个人的分立与分配正义): 330, 133, 498 and 460 (both on 133)

Setiya, Kieran (克尔兰·赛提娅): xlvi

Share of the Total View about how we can do the most good (有关我们如何做最有益事情的总量平分观): 253

Shipwrecks (船舶事故情形), 救自己还是救陌生人: 139–41

Sidgwick (西季威克): 与康德的比较, xxxiii–xxxiv; 他的伟大, 我们为什么应该读他的著作, xxxiv–xl; 实践理性的二元论以及“最深刻的问题”, 6–7, 130–49, 498 (on 133) 499 (on 143) 461 (on 143); 两视角论证: 134–6; 西季威克的后果论是作为道德的外在对手, 168–9, 171, 500; 提醒要小心隐藏的同义反复, 234; 西季威克与自然主义谬误, 465; 接受行动后果论如何可能是间接地自我挫败的, 使得事情之进展没有本该有的那么好, 251–2, 404–6; 提出但是拒绝了基于理想选择的福祉论说, 496–7; 论快乐的强度, 460; 看起来是一些错误的事情, 453

significantly non-optific principles (有意义的非最优原则): 380, 399–400, 477

Smith, Michael (迈克尔·史密斯): 79–80, 482

state-given reasons (状态给予的理由): 50–1, 420–32, 442–4

Stoics (斯多葛学派): xliii, 371, 453, 513

Subjectivism about reasons (理由的主观主义): 参见 Reasons (理由)

substantive normative beliefs (实质性的规范信念): 70–3, 456 (on 70), 62, 105–6, 125, 247

substantively rational (实质上合理的): 62–3, 78, 121

Suffering (受苦), 参见 pain (痛苦)

sufficient reasons (充分理由): 32–3 (有些人在不同的意义使用“充分的”, 意思是“决定性的”)

Suicide (自杀): xliii, 198, 200, 235, 239, 453

Surgery (手术情形), 有证成的强制: 196–7

Tautologies (同义反复), 参见 concealed tautologies (隐藏的同义反复)

teleological theories (目的论理论): 237, 246–7; 参见 Consequentialism (后果论)

Telic (目的式): desires (目的式欲望), 44; reasons (目的式理由), 52

Temkin, Larry (拉里特姆金): xlvi, 498

temporally-neutral desire-based sense of 'best for' (“对……是最好的” 之时间中立且基于欲望的意义) : 105

thin theory of the good, Rawls's (罗尔斯的稀薄善理论) : 103-5

Thomist view (托马斯主义观点) : 158

thought-experiments (思想实验) : 爱因斯坦与黄金规则, 329; 285, 328, 350, 355-6, 382, 476

Threshold Objection (门槛反驳) : 308-12

time, attitudes towards (对时间的态度) : 我们对于眼下的实际偏向, 对于来年的虚构偏向, 对于未来周二的漠不关心, 46, 56-7; 我们对于未来的偏向, 197; 时间中立性, 495 (on 57)

Transitivity (传递性) : 128, 459

Transplant, Act Utilitarianism and Scanlonian Contractualism (移植情形, 行动效用主义和斯坎伦式契约论) : 363-5, 473 (on 363)

treating someone as a means, merely as a means, and as a mere thing (把某人作为手段来对待, 把某人纯粹当作手段来对待, 把某人当作物件来对待) : 10-12, 212-28, 463-4 (on 212-29); 作为手段的伤害, 228-232; 与斯坎伦式契约论, 360-6

Triple Theory (三重理论) : 411-17; 以及卷二

trivial benefits and burdens (微不足道的利益与负担) : 307; 以及卷二

Tunnel (隧道情形) : 218; 我们不杀人的义务优先于我们救人的义务, 对此信念的一个反例, 450; 与桥梁情形的比较, 219-220, 228-30, 364-6, 449-51

Two Viewpoints Argument (两视角论证) : 134-6

UA-optimific principles (UA 最优原则) : 377-9, 404-7

UF-optimific principles (UF 最优原则) : 405-7, 482-3 (on 406)

Unanimity Condition, for the Consent Principle (同意原则的全体一致条件) : 188

uniqueness condition, for the Kantian Contractualist Formula (康德式契约论公式的独特性条件) : 358

universal acceptance (普遍接受) : 285-6, 341-2

universal compliance, or some principle's being universally followed (普遍服从或某个原则得到普遍的遵循) : 343, 405-7, 482-3 (on 406)

Universal Law (普遍法则), 参见 Kant's Formulas (康德的各种公式)

Unjust Punishment, and the Rarity Objection to Kant's Formulas (不公正的惩罚情形, 以及对康德的公式的稀罕性反驳) : 330-1

unreasonable in Scanlon's intended sense (在斯坎伦想要的意义上的“不合情理”) : 360-5, 416, 473

Utilitarianism (效用主义) : 373; 其享乐主义形式, 246, 更好地重述为有关痛苦之主张, 463; 作为不偏不倚理由之行动后果论的一种形式, 它也许可以称为是道德的

外部对手, 168-9, 500; 康德有关享乐主义规则效用主义的主张, 255-6, 410; 罗尔斯提出的替代选项: 349-55; 与斯坎伦式契约论, 362-4, 473 (on 364); 以及参见 Consequentialism (后果论)

valid consent (有效同意): 195-200

Values (价值): 要被尊重的: 236-9, 243, 478; 以及参见 Good (and bad) (好与(坏))

value of life, two views about (有关生命价值的两种观点): 238-9

veil of ignorance (无知之幕): 参见 Contractualism, Rawlsian (罗尔斯式契约论)

Veto Principle (否决权原则): 192-4

virtue ethics (德性伦理学), 其系统的形式与动机后果论: 375, 475

Weak Moral Rationalism (弱道德理性主义): 144

Well-being (福祉): 39; 各种福祉理论: 享乐主义的, 实质善的, 基于欲望的, 39-40, 74; 在蕴含理由的意义上什么是对某个人最好的, 102, 在基于当前选择的、享乐主义的且是时间中立意义上的, 103-6; 罗尔斯的稀薄善理论, 103; 是否存在基于欲望的自利理由, 102

What if everyone did that? (如果每个人那么做会怎么样?): 17-18, 286, 301-320

What if everyone thought like you? (如果每个人像你那样想会怎么样?): 286, 320, 471

what someone is doing, his doing of it (某人在做什么, 其做法): 290

what someone is intentionally doing (某个人正想做什么): 297

Whimsical Despot (反复无常的专制者), 对理由的非自愿回应 47; 以及欲望的合理性, 125-6; 以及状态给予的理由, 420-30; 以及主观主义, 457; 以及对康德式契约论公式的一种反驳, 476 (Rosen's malicious gremlin (罗森之恶意的精灵))

Whole Scheme View about how we can do the most good (我们如何做最有益的事情的整体框架观): 254-6

Wide Dualism (广义二元论): 140-1

wide value-based objective views of rationality and reasons (“合理性和理由”之广义的基于价值的客观观点): 137-41, 186, 460 (on 137); 以及参见 Reasons (and Rationality) (理由(和合理性))

Williams, Bernard (伯纳德·威廉斯): 受到西季威克有关性道德评论的误导, xxxviii, 452-3; 论主观主义, 65, 77; “生命要有意义”, 387; 论行动者遗憾, 461; 论摩尔 465

Williams T. C (T. C. 威廉斯): 339

willing as a universal law (意愿作为普遍法则): 285-8, 301-2

Wood, Allen (阿伦·武德): 论康德的人性公式与尊重人, 233; 论性、自杀与撒谎, 235; 论康德有关价值的主张, 235, 241; “即使是最坏的人也有尊严”, 240; 谴责康德的普遍法则公式, 293; 以及卷二

Wrong (不当): “不当”的意义与各种不当性: 150–174; 事实相对的、证据相对的和信念相对的意义, 151–164, 461; 道德信念相对的意义, 151, 171; 不可定义的意义, “绝不可做”, 165–6, 169–70, 173, 451; 其他可定义的意义, 164–74; 可用来定义“应当”和“正当”以及“道德上许可的”, 165; 不当性本身以及“使行动不当”的那种属性, 368–70

Wrong-Making Features Objection (致不当特征反驳): 390–98, 448–51

《论重要之事》

全三卷简介

卷一 论重要之事

第一部分 理由

第二部分 原则

第三部分 理论

在本卷，帕菲特主要论证存在客观理由或价值，而否认这一点则会导致道德虚无主义。道德主观主义者与相对主义者表面上否认存在着客观的价值与理由，但实际上只是没有搞清楚自己的真实立场。基于这种立场，帕菲特重新探讨当今最重要的三种规范伦理学理论（后果主义、康德式义务论和斯坎伦的契约论），认为这三种理论之所以表现得水火不容，实际上是缘于错误的主观主义。帕菲特利用各种各样的思想试验来论证：一旦消除三种理论之中的主观主义成分，做出合理的重构，那么它们是趋同的。三种理论之间的所谓根本分歧不过是人们从不同侧面攀登同一座山峰而所见到的不同风景，就其最终目的而言殊途同归。

卷二 挑战与回应

第四部分 评论

第五部分 回应

本卷的第一个部分是呈现当今四位一流的道德哲学家对于帕菲特登山论的挑战与质疑，另一个部分则是帕菲特对于这些质疑与挑战的回应。主要的质疑有二：其一是帕菲特所谈的康德式伦理学与斯坎伦的契约论不再是其原初理论，其二是这种重构会失掉原有理论最有价值的成分。阿伦·武德则认为康德伦理学最重要的原则并非是普遍同意原则。而赫尔曼则是从方法论入手，认为帕菲特的论证方法是杂合的，但这绝不是康德式伦理学的特点。斯坎伦认为帕菲特曲解了自己的契约论。帕菲特的回应主要是通过更多的思想试验表明，他基于客观理由的重构能够更好地解释我们的道德直觉，他偏离那些理论的原型只是为了更好地修正它们，取得道德理论的发展与进步。

卷三 元伦理学根基

第六部分 规范性

本卷是帕菲特试图为其规范伦理学寻求元伦理学的根基。与规范伦理学领域寻求共识形成鲜明对照的是，帕菲特在元伦理学领域论证其他的立场是错误的，似乎带有独断论的特色。然而，帕菲特的整个目的是表明客观理由、客观价值是存在的，由此存在着客观的道德规范。凡属与这种立场相冲突的成分，均是需要放弃的，比如康德的实践理性观就是被拒绝的。帕菲特对各种元伦理学理论的处理，在这个意义上讲与对规范伦理学的处理是一样的。因为一切非认知主义均无法与客观的价值与理由相容，所以是要被否定的。不仅如此，帕菲特认为，一切自然主义的认知主义都会放弃道德的规范性，由此也是无法得到辩护的。同时，非自然主义的形而上学认知主义就会走向摩尔的立场，导致柏拉图意义上的道德实体，这又得不到现代科学理论的支持。由此，帕菲特只剩下一种立场：非自然主义且非形而上的认知主义。而与这种认知主义可相提并论的类型就是数学与逻辑。这就是卷三中的全部论证所试图推出的。

德里克·帕菲特是当代西方伦理学界最重要、最有原创性的一位思想家，他在1984年出版的《理与人》被认为是自密尔的《功利主义》和西季威克的《伦理学方法》以来功利主义伦理学领域的一座里程碑，开创了一系列至今仍在热烈讨论的议题。在两卷本的《论重要之事》中，帕菲特倾其一生的功力试图表明他能够在三种被认为截然不同的规范伦理学理论——功利主义（或后果主义）、康德伦理学和契约论伦理学——之间实现一种调和，并把他的论证建立在他对元伦理学和规范伦理学的全面考察和深入分析的基础上。因此，这部重要论著不仅在某种程度上挑战了人们对规范伦理学乃至道德的本质和功能的传统理解，也把规范伦理学和元伦理学的联系提升到了一个新的水平。帕菲特在本书中提出的论点和论证并非无懈可击，但也正是这个缘故，就像他的早期著作一样，本书也为伦理研究设置了一个新的制高点。我相信本书中译本的出版将会推进我们对西方伦理学的认识和研究。

徐向东，浙江大学人文学院哲学系教授

帕菲特是一位绝对重量级的，尤其是在令人尊敬的同行之间享有顶级声誉的当代哲学家。《论重要之事》不但是帕菲特几十年哲学生涯中的里程碑式作品，而且在某种程度上构成了晚近几十年世界哲学范围内的实践理性转向的总结。在为一种精致的更有包容力的后果论做出当代最高水准论证的同时，此书也提供了自西季威克以来对道德哲学谱系的一次最辉煌的重构。另外，此书译笔慎重谨严，不但可以作为研究性工作的有力参照，而且将为淬炼提高汉语言的论证精密度做出贡献。

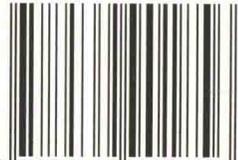
应奇，浙江大学外国哲学研究所教授

在被誉为“近乎天才之作”的《理与人》（1984年）出版之后，帕菲特已经奠定了他在道德哲学领域中的大师地位。《论重要之事》（2011年）曾是哲学界的一个神秘传说，手稿在同行之间广泛流传和讨论，长达十多年之久。如今，葛四友教授和阮航教授精工细作的译本为中文读者呈现了这部令人期待已久的巨著，让我们可以与当代杰出的哲学家一起来思考道德判断的客观性这一核心问题。

刘擎，华东师范大学政治学教授

上架建议：哲学

ISBN 978-7-5699-0185-6



9 787569 901856