诺布效应视野下的陪审团公正性

李锋锋

[摘要] 实验哲学家约书亚·诺布提出的诺布效应显示，与一个好的副作用相比，人们更加倾向于认为坏的副作是行为者有意而为之的，据此，人们更倾向于对造成坏的副作用的行为者进行惩罚，但不会因为是好的副作用而表扬行为者。这就造成了认知判断和责任归因的非对称性的状况。当这种现象出现在陪审团中，就会对嫌疑人本应得到的公平、陪审团本应维护的正义以及陪审团成员无偏见地进行判断的能力构成伤害。

[关键词] 实验哲学；约书亚·诺布；诺布效应；陪审团；责任
[中图分类号] DF0－054 **[文献标识码]** A **[文章编号]** 1004－0633（2013）01－087－05

实验哲学家约书亚·诺布（Joshua Knobe）通过哲学实验研究发现：与道德上好的副作用相比，人们更倾向于将道德上坏的副作用看作是有意而为的，这就是诺布效应。托马斯·纳德霍夫尔（Thomas Nadelhoffer）通过实验也证实了这种观点：人们对有意行为的归因经常受评价性考量的影响。果真如此，我们就有理由怀疑，陪审团能否公正地、无偏见地做出关于有意行为的判定。如果陪审团也无例外会受诺布效应的影响，那么，陪审团成员就无法实现无偏见地断定嫌疑人的动机和意图了，而这与陪审团设立的目标是相违背的。嫌疑人的行为动机或者意图是构成犯罪成立的重要主观因素，需要陪审团公正地对其进行断定。陪审团的判断结果决定嫌疑人是否被起诉、是否有罪，决定嫌疑人的命运。因此，陪审团无偏见地、公正地进行判定，不仅是对嫌疑人的负责、也是对社会正义的负责。

一、诺布效应的提出及其普遍性

诺布在一系列的哲学实验中发现，人们断定一个副作用的产生是不是行为者有意而为之的，依赖于这个副作用的道德性质，副作用是好的还是坏的将决定着人们对其是否是有意的判断。诺布做了两个实验。

实验一的78名受试者来自曼哈顿公园，每名受试者被随机地分到“危害条件”组或者“帮助条件”组。危害条件组的受试者读到以下场景：

某公司的副总裁对董事长说：“我们正在考虑进行一个新的项目，它会为我们增加利润，但同时会危害环境。”

董事长答道：“我才不关心什么环境危害，我只想尽我所能赚取更多的利润，现在就开始这个新项目吧。”

他们就开始了这个新项目，果然，环境被破坏了。[1]

然后，要求受试者判断，董事长要为他的所做担负多大的责任（范围为0－6），并说出他们是否认为懂事长是有意在破坏环境的。结果显示，有82%的受试者认为董事长是有意破坏环境。[2]

帮助条件组的受试者读到基本相同的场景，只是将“破坏”一词改为“帮助”：

某公司的副总裁对董事长说：“我们正在考虑进行一个新的项目，它会为我们增加利润，同时也会帮助环境。”

* 感谢导师曹剑波教授对此文给予的细心指导！
[收稿日期] 2012－09－05
[作者单位] 李锋锋，厦门大学哲学系。 福建厦门 361005

董事长答道："我才不关心什么帮助环境，我只想尽我所能赚取更多的利润，现在就开始这个新项目吧。"

他们就开始了这个新项目，果真，帮助了环境。[3]

然后，要求受试者判断，董事长应得到多大程度的赞扬（范围为 0－6），以及他们是否认为董事长是有意帮助环境。结果显示，有 77% 的受试者认为董事长不是有意在帮助环境。[4] 与上面的 82% 相比，这是截然相反的结果，通过卡方分布分析得到 X^2（1，N＝78）＝27.2，p＜0.001。[5] 差异性是非常明显的。

为了验证结论的可靠性，诺布做了另外一个实验。

实验二的 42 名受试者也来自曼哈顿公园，每一名受试者被随机地分到"危害条件"组或者"帮助条件"组。危害条件组的人们读到以下场景：

中尉正和一名中士谈话，并下命令说："把你那个班派到汤普森山顶。"

中士说："如果我派我的班到汤普森山顶，我们就直接将战士们推到了敌人的火力线前沿，他们中的有人一定会被射杀。"

中尉回答说："听着！我知道他们会在敌人的火力线前，而且他们中的有人也会被射杀。但是，我一点都不在乎发生在战士身上的事情，我所在乎的是能够控制汤普森山！"

这个班被派到了汤普森山顶，不出所料，战士们进到了敌人的火力线前，而且他们中的有人被射杀了。[6]

然后要求受试者判断，中尉要为他所做的担负多大的责任，并且，他是否是有意让战士们进到敌人的火力线前。

在帮助条件组的受试者有一个同上基本相同的情境描述：

中尉正和一名中士谈话，并下命令说："把你那个班派到汤普森山顶。"

中士说："如果我派我的班到汤普森山顶，我们就把战士们带离了敌人的火力线，他们就会得救。"

中尉回答说："听着！我知道会把他们带离了敌人的火力线，而且我也知道，如果不这样的话，他们中的有人会被射杀。但是，我一点都不在乎发生在战士身上的事情，我所在乎的是能够控制汤普森山！"

这个班被派到了汤普森山顶，不出预料，战士们脱离了敌人的火力线，他们因此幸免于难。[7]

然后要求受试者判断，中尉应为他所做的得到多大程度的表扬，以及他是否是有意将战士们带离敌人的火力线。

试验再次出现了两个截然相反的结果：在危害条件下，77% 的受试者认为这个副作用是中尉有意而为；然而，在帮助条件下，有 70% 的受试者认为中尉不是有意的。[8]

在这两个实验中，总共参与的受试者有 120 人，在危害条件中，认为需要承担的责任的程度在 0 至 6 这个范围中平均达到 4.8；但是在帮助条件中应得到表扬的程度仅为 1.4。通过以上的哲学实验，诺布认为存在一种不对称，即人们更加倾向于因为一个坏的副作用而惩罚行为者，而不是为了一个好的副作用而去表扬行为者。诺布说："人们对于做出惩罚和表扬决定的不对称是基于对'有意的'（intentional）这个概念应用的不对称性，即与一个好的副作用相比，人们更加倾向于去说一个坏的副作用是有意做出的。"[9]

诺布效应在责任归因中广泛的存在。如第一，在现今非常紧张的医患关系中，一般说来，一个医生对待他的病人都受职业道德约束，那就是在自己的能力之下治好病人的病，这一点是不用怀疑的。但是，任何病症的治疗都有风险，特别是重大手术中，这和人们普遍认为的"是药七分毒"是一样的。在旁观者特别是病患家属看来，当一个医生用一种治疗方案治好病的时候，他们认为这是医生应尽的职责；但是，当医生没有治好病，甚至是一段时间的治疗后还出现了恶化的迹象，那么，他们就会认为这是医生的行为所致，医生就要承担责任了。显然，这是诺布效应在我们生活中的具体体现。第二，在股市中也存在着明显的这样的责任归因心理和状况。一个人依据自己的股市知识和经验选择了两只股票，当其中的一只股票上涨了，人们认为他有一个好的运气；当另一只股票下跌时，人们会认为他自己没有选好股票或者没有看准时机等，他的知识和经验是有问题的，这是他自己的原因。第三，这里要说到与下文将要举的例子类似的一个状况就是，一名警察在追捕一名逃犯的过程中，不小心使得枪走火了，当子弹恰巧伤到另外一名正在通缉的罪犯的时候，人们普遍认为这名警察不负有责任，还要受到褒奖；而当子弹伤到一名群众的时候，人们就认为这是警察的责任了。其实，生活中有很多的这种状况，结果是好的没有什么可表扬的，一旦出现了坏结果，行为者就必须承担责

任，用北方人们日常中的话来说就是：好了好了，没事算你幸运，要是有什么事就让你吃不了兜着走。这句话就非常典型地体现了此种心理。所以，在人们的日常生活中，诺布效应的存在是普遍的，它影响着人们的责任归因心理。

二、诺布效应对陪审团公正性的影响

在罪行特别是刑事罪名成立的条件中，主观上的故意（直接或者间接）是其基本条件。而对主观上是否是有意或故意的断定，将直接影响着案件审理的进展和罪名的成立，所以，这在案件调查和审理中是极为重要的步骤。接下来，本文将从诺布效应着手，论述陪审团所受到的影响将影响到陪审团是否能实现公正的目标。

根据诺布效应，可以看到人们对行为善恶的判断影响他们对这个行为是不是有意做出的判断，并依据这种有意性的判断来确定当事人是不是要为这件事负责。然而，由于行为善恶对行为是否有意作出的判断的影响具有非对称性，恶的影响更大，这时我们就产生了一种担心，即陪审员在一个道德上恶的事件中不能公正地断定行为者是否是有意做出行为。而且这种担心是有实验依据的。

为了更好地弄清楚副作用的道德善恶是否影响陪审团的公正性这个问题，纳德霍夫尔基于这个事件原型设置了一个情境，并有126名大学生参与试验。其中的一半学生看到这样的情境：

> 案例1：想象一个贼开着一辆载满才偷来的赃物的车。当他在等红灯的时候，一个警察手中挥舞着一把枪来到了车窗前。看到这个警察，小偷加速开过了十字路口。令人惊讶的是，当车子加速的时候，这名警察竟然抓住了车子的一边。为了能逃脱——虽然清楚地知道这样做会置警察于危险中，小偷突然转弯走之字形路线。但是小偷并不在意，他就是想要逃跑。这位警察很不幸，小偷想要把他甩开的尝试成功了。结果，这个警察滚进了来往的车流之中，受了致命伤。几分钟后他便死了。[10]

这一半参与者需要回答以下问题：第一，这个贼明知这样会引起警察的死亡吗；第二，这个贼是有意造成警察死亡的吗；第三，这个贼应为警察的死负有多大的责任（范围为0－6，0是没有责任，6是负有很大责任）。他们回答的结果如下：第一，75%的参与者认为明知会引起警察的死亡；第二，37%的参与者认为是有意造成警察死亡的；第三，在6个基点下，平均责任评定值为5.11。[11]很明显，人们（75%）更加倾向于认为这个贼对于警察的死亡是明知的，也可以说（虽然是37%）倾向于认为是有意为之的。

另外，为了弄清楚是警察死亡的恶还是感知到的小偷在道德上的罪责，或者两者都广泛影响着参与者对认识与意图的归因，纳德霍夫尔给了另外一半参与者在结构上与第一个情境相同的例子，只不过这次是一个无辜的司机导致了一个行为未遂的劫车者的死亡，具体情境如下：

> 案例2：想象一个男人在车里等红灯。突然，一个盗车贼靠近了他的窗口，手中还摇晃着一把枪。当看到这个贼时，司机惊慌不已，加速驶离十字路口。令人惊讶的是，在车加速的时候，这个盗车贼居然抓住了车的一边。司机为了能逃脱——同时清楚地知道这样做会置盗车贼于危险中，他突然转弯走之字形路线。但是这个司机并不在意，他只是想要逃脱。盗车贼很不幸，司机想要把他甩开的尝试成功了。结果，这个盗车贼滚进了来往的车流之中，受了致命伤。几分钟后他便死了。[12]

然后，参与者被要求回答与第一个案例中相同的三个问题，即第一，司机明知这样会引起盗车贼的死亡吗；第二，司机是有意造成盗车贼死亡的吗；第三，这个司机应为盗车贼的死负有多大的责任（范围为0－6，0是没有责任，6是负有很大责任）。这一半的参与者给出的答案是：第一，51%的参与者认为是明知会引起盗车贼的死亡；第二，10%的参与者认为是有意造成盗车贼死亡的；第三，在6个基点下，平均责任评定值为2.01。[13]与第一个案例相比，这三个问题答案的差异性是非常明显的，对于第一、二问题的回答经过卡方分布分析分别得出 X^2（1，N＝126）＝7.62，$p<0.01$；X^2（1，N＝126）＝12.94，$p<0.001$，[14]同时，两组参与者对于责任认定也是存在巨大差异的，分别为5.11和2.01，所以，纳德霍夫尔认为，“就最初证据看来，道德考量的确解释了参与者判断的不对称性”。[15]

这个试验结果和诺布效应显示的结果是一致的，初步说明了陪审员的判定是受到道德考量的影响，同时，关于受害者的道德性的考量也影响到了对被告的行为主观性的判定。因为，当被告人是贼而被害者是警察的时候，人们更倾向于认定被告人是明知（75%）一个坏结果的发生，进而判定这个行为是有意的（37%），从而结果就是要承担更大的责任（5.11）；但当被告人是无辜者而被害者是贼的时候，人们对明知性（51%）和有意行为（10%）的认定就不是那么比较倾向了，从而对于

责任的认定也没有那么严重了（2.01）。可见，不仅是副作用的恶，还有行为者应担负的道德责任（案例 1 中的贼和案例 2 中无辜的司机所承担的道德责任是不同的）影响了对于有意行为的判定。为了佐证这种观点，纳德霍夫尔又引出了戴孟德（Desmond）案件、海雅姆（Hyam）案件及雷吉纳状告坎宁翰的案件，通过结果的分析，我们可以更加坚定地认为，对副作用的恶与承担的道德责任的考量带来了对于行为有意的判定的偏见性影响，从而影响到了对责任的认定的判断。这说明了，在诺布效应下，陪审员公正性地做出判定的能力被破坏了。

另外，根据艾利克的“有罪控制的责任模型”（CCMB）我们也可以看到，“个人控制的判断和责任归因会受到相对无意识的、自发性的对于精神的、行为的和结果等因素的评价的影响，这种自发性的评价是对涉及到的危害性事件和人的情感性反应”，[16]这个反应不仅能被诸如个人控制的证据性因素触发，而且还能被如一个人的外表、名声、社会地位等证据外因素的触发。因此，艾利克认为，“关于个人控制的判断——因此也是对于应承担的道德责任——就会不自觉地受涉及到的事件和行为的情感性反应的影响”，这种自发的责任预设导致了陪审员有选择性地寻找那种能够支持罪责认定的证据，同时忽视那些有可能减轻或开脱责任或罪行的证据。[17]正如上文中提到的偷车贼案件中，如果分析正确的话，那么在偷车贼受审之前，陪审团那里就已经预定了他的有意行为及责任。就是说，“行为的不道德性可以自动触发陪审员进入责任认定的缺省模式——这个模式使得他们受到负面的、相对而言无意识的反应的影响，使得他们对罪行的评判与对使被告有罪的结构性关联的评价均带上了偏见”，这些自发的责任确定偏见并不是“理性标准的例外”而是“责任认定的内在特点”。[18]这就更说明了要是想消除这些偏见性因素的影响是相当困难的。其实，影响人们判定的因素不仅是这些，还有预见性，① 文化差异性与结果严重度的存在。②

所以，我们可以清楚地看到，陪审团成员要公正地对是不是有意行为作出判定是不容易的，因为影响因素是人们的道德考量，并且，这种影响是无意识的。其实，道德是一个人出生下来之后就被不断灌输的“软约束力”，它在社会当中特别是日常生活中的规范作用要比法律的约束力更为基础、内在和及时，所以，要想使陪审员摆脱这种道德考量而进行判定是不可能的。但是，我们应该努力消除掉影响公正性判定的偏见性因素，使得公正性得到最大程度的实现。

三、对保证陪审团公正性的思考

对如何消除偏见性因素，上文已经提到了其本身所具有的“内在性”特点，所以，这就增加了困难度。但对于这个问题的思考，思想家们仍不懈努力，有许多学者已经提出了一些解决之道。在马勒与尼尔森看来，“关于有意性或者意图的判断对裁决和量刑有着重大的影响，并不仅仅是预示它们，所以，应该尽我们所能将意图判断与情感性评价或者罪责的担负分离开。”[19]根据上文的论述，我们所要做的工作，就是使对有意行为的判定与道德考量因素和情感性评价等区分开来，只有这样，陪审员才能公正地做出判定。现在问题是，该如何实施这种设想。马勒与尼尔森给出的解决之道就是“劝诫陪审员将评价性情感放到一边”[20]，但是，当我们告诉陪审员当他们在判定有意行为时是会受到道德考量的因素或者评价性因素的影响，那么，他们是否在判定过程中就力图消除偏见呢？我们并不这样认为。也许，他们找到一些理由说就应该判定某一行为是有意为之的，当这种影响因素被表明后，就又回到了艾利克的 CCMB 中讲到的观点了：当自发性评价触发后，陪审员可能会有选择性地寻找那种能够支持罪责认定的证据，同时忽视那些有可能减轻或开脱责任或罪行的证据。证据表明，“抑制带有偏见反应的行为反而会增加此偏见出现的频率”，[21]如此来就事与愿违了。从认知心理学上来说，这种方法也不会成功的。艾利克森（Erikson）和西蒙（Simon），雅各比（Jacoby）、林赛（Lindsay）和托特（Toth），凯尔斯壮（Kihltstrom）等人认为，我们的认知过程是不可能

① David A. Lagnado 和 Shelley Channon 在文章“Judgments of cause and blame: The effects of intentionality and foreseeability”（Cognition 108（2008）：754－770.）认为有意行为中的预见性具有强大的影响，当一个行为具有很强预见性的时候，那么，它就具有与这个强度相应的因果联系，行为者因此更应受到处罚。

② Philip E. Tetlock，William T. Self，and Ramadhar Singh 在文章“The punitiveness paradox: When is external pressure exculpatory － And when a signal just to spread blame?”（Journal of Experimental Social Psychology 46（2010）388－395.）中通过对美国人和新加坡人进行的实验研究发现，在责任归因上，相较于美国，新加坡更会受到结果严重度的影响；同时，美国会随着增加的可减轻处罚条件而减轻处罚，但是新加坡不会受到这个因素的影响等。

被有意识地处理的。虽然人们认识到了认知偏见的存在及其重要性，但是，他们随后对其思想和感觉的控制能力是非常有限的。①

那么，我们是不是可以跳出这个圈子外再回头看看。之所以力图消除偏见性因素的影响，就在于它构成了一个更为重要的结果，那就是影响了对于有意行为的判定，进而关系到司法行为中行为者或者被告人的判决结果。如果对于有意行为的判定并不决定着这个行为者或被告人的罪责成立，而只是影响因素之一，与其他诸如预见性、技术性②、事实性、以及文化差异性与结果严重度等一块构成决定性因素，这样就不致使将焦点聚于一个因素上面，通过把这一个因素的“解构”来实现对其的解决。

其次，我们同时也要关注信念形成的过程。陪审团具有的这种偏见性也是人们认知的偏见，上述我们提到了它对于嫌疑人与社会正义构成的危害，但是，我们不能否认其具有的内在性特点，不能否认作为认知的偏见，它的形成与信念的形成具有同质性，也就是说，一个信念的形成构成了人们对一件事情的初步判断，这个信念构成了人们的认知基础（其实也是知识形成的基础），在此基础上产生的判断不可避免地带有了信念的偏见，所以，对于偏见的解决，我们也需要着手解决信念形成过程中存在的偏见。

最后，我们将借以纳德霍夫尔的话来结束本文：“直至我们对这个问题的本性与深度都有一个更好的理解前，我们将无法提出什么可行的解决措施。涉及到的重要一步就是要更加关注大众心理学和大众道德之间的关系。还有一点就是更仔细地考察犯罪意图这个概念在日常语言与刑法中所扮演的角色。这是一项需要哲学家、心理学家和法学家携手工作的调查研究。如果我的工作促进了沿着这条道路进行进一步研究的话，那么它就是成功的，即使我留下了一些重要的没有回答的问题。”〔22〕

【参考文献】

〔1〕〔2〕〔3〕〔4〕〔5〕〔6〕〔7〕〔8〕〔9〕Knobe，J.：Intentional action and side effects in ordinary language. Analysis 2003，63：p191，p192，p191，p192，p192，p192，p192－193，p193，p193.

〔10〕〔11〕〔12〕〔13〕〔14〕〔15〕〔22〕Nadelhoffer，T.：Bad acts，blameworthy agents，and intentional actions：Some problems for juror impartiality. Philosophical Explorations Vol. 9，No. 2，June 2006. P209，p209，p209－210，p209－210，p210，p210，p215－216.

〔16〕〔18〕Alicke，M. D.：Culpable control and the psychology of blame. Psychological Bulletin. Vol. 126，2000. No. 4：p558.

〔17〕Nadelhoffer，T.：Blame，badness，and intentional action：A reply to Knobe and Mendlow. Journal of Theoretical and Philosophical Psychology. 2004，24：p262.

〔19〕〔20〕Malle，B.，and S. Nelson.：Judging mens rea：The tension between folk concepts and legal concepts of intentionality. Behavioral Sciences and the Law. 2003. 21：p576.

〔21〕Wilson，T. D.，and N. Brekke.：Mental contamination and mental correction of unwanted influences on judgments and evaluations. Psychological Bulletin 1994. 116（1）：p133.

（责任编辑：谢莲碧）

① 这里可以参阅的文章有：Bargh，J. A. 的“Conditional automaticity：Varieties of automatic influence in social perception and cognition”和 Logan，G. D. 的“Automaticity and cognitive control”（In Unintended thought：Limits of awareness，intention，and control，edited by J. S. Uleman and J. A. Bargh. 1989. New York：Guilford Press.）以及 Wegner，D. M. 的“White bears and other unwanted thoughts”（1989. New York：Viking Press.）和“You can't always think what you want：Problems in the suppression of unwanted thoughts”（1992. In Advances in experimental social psychology. Vol. 25，edited by M. P. Zanna. San Diego，Calif.：Academic Press.）

② 关于技术性在责任归因中的影响，Mele，A. 和 Moser，P.（Intentional action. Nous，28，39－68. Reprinted in A. Mele（Ed.），（1997）. The philosophy of action. New York：Oxford University Press.）认为这是必要的，但 Nadelhoffer 并不这样认为，他在文章“Skill，luck，and intentional action”（Philosophical Psychology 18：343－354）中对此进行了反驳。