

道徳的責任の文脈において 無知と潜在的態度は類比的に論じうるのか¹

薄 井 尚 樹

要旨：道徳的責任の文脈において無知が正当な弁明とならないことがある。この「責められるべき無知」の構造を明らかにすることは、道徳的責任の認識条件をめぐる重要な問題のひとつである。本稿はこの問いをめぐる議論のなかに「潜在的態度」と呼ばれる心的状態がどのように位置づけられうるのかを考察する。その考察を通じて、「気づきの欠如」と「過去の浅はかな行為」というふたつの論点が、潜在的態度と無知を類比的に捉えることに関わっていることを明らかにし、それにもとづいて潜在的態度の責任をめぐるなされてきた哲学的議論を分析する。

1. はじめに

たとえ不正な行為であっても、それがあることに気づかずになされた行為であった場合、その行為者が非難に値しない、つまり道徳的責任を免れることはよくある。Rosen (2003) の例を借りて、あるひとが森のなかで道に迷ってしまい、なんの標識もなかったために私有地に入り込んでしまったところを想像してみよう (Rosen, 2003, p. 62)。不法侵入は不正な行為であり、ふつうその行為者は非難に値するだろう。しかしいまの場合はそう主張できないように思われる。たしかにこの場合でも、問題の行為そのものが不正であることは間違いない。しかしそのひとは、ここが私有地だと自分は気づいていなかったのだと述べるだろうし、そのことはたしかに、正当な弁明となりうるように思われる。このように、無知（気づきの欠如）が弁明となって、行為者はときに道徳的責任を免れることがある。

とはいえ、いつでもそうであるわけではない。もうひとつの事例として、ある医師が患者に薬を処方するところを考えてみよう。この医師は最新の研究を参照していなかったために、安全な代替薬の存在に気づかずに、副作用がともなう薬の処方を選択し、結果として患者に深刻な障害が生じたとしてみよう（このような事例はたとえば Wieland (2017) で言及されている）。この場合、たとえそれが無知からなされた行為だったとしても、その医師は道徳的責任を免れないかもしれない。したがって道徳的責任の文脈において無知はつねに正当な弁明となるわけではない。それでは、無知はどのようなかたちで道徳的責任にかかわるのだろうか。これは道徳的責任の認識条件 epistemic condition をめぐる重要な問いのひとつである。

本稿の目的は、この問いをめぐる議論のなかに潜在的態度 implicit attitude という心的状態がどのように位置づけられうるのかを論じることにある。そのさい、道徳的責任という文脈のなかで無知と潜在的態度を類比的に論じることによってどのような論点がともなうのかを明らかにし、そのうえで、潜在的態度（からなされる行為）の責任についてなされてきた議論を分析する。

本稿は以下のように進行する。まず、無知からなされる行為と道徳的責任の関係を描写する基本的な図式を概観し（第2節）、つぎに潜在的態度を測定することを可能にした代表的な実

験手法を紹介する（第3節）。最後に、道徳的責任の文脈において無知と潜在的態度とのあいだに類比的な議論が成立するために必要な論点を明確にする（第4節）。

2. 無知からなされる行為の構造

先述の医師が、安全な代替薬の存在に気づいていなかったにもかかわらず、患者を傷つけたことについて道徳的責任を負わされるのはなぜだろうか。つまりその無知が責められるべき無知 culpable ignorance であるのはなぜだろうか。Smith（1983）は、責められるべき無知に共通する構造をこう説明する。

〔引用者註：責められるべき無知に〕関連するどの事例にも、ある一連の行為がともなう。その最初の行為において行為者は、自分の認知的な立場を改善しない（あるいは積極的にそれを損なう）。そしてその結果として生じる無知のために、その後の行為において行為者は不正をするのである（Smith, 1983, p. 547）

Smith にならって最初の行為を「浅はかな行為 benighting act」（Smith, 1983, p. 547）と呼ぶことにしよう。たとえば先の医師は、最新の研究を参照しなかったことで自身の医療的な知識をアップデートしなかった。これが浅はかな行為に該当する。Smith によると、この浅はかな行為が非難に値するとき、その結果として生じる無知は責められるべきものとなる（Smith, 1983, p. 548）。したがって、ある無知が責められるべきものである（それゆえそこから生じる行為が非難に値する）のは、過去の浅はかな行為が非難に値する場合にかぎられることになる。

責められるべき無知は、安全な代替薬の存在に気づいていないことのように、事実にかかわる無知にかぎらない。Rosen（2003）は、事実的な無知について Smith が指摘する構造が、道徳的な無知にもあてはまると主張し、それを「等価テーゼ the parity thesis」（Rosen, 2003, p. 64）と呼ぶ。たとえば、人種差別が不正であることに気づかないまま、ある特定の人種に属する人々に差別行為をするひとを考えてみよう。等価テーゼによれば、問題の差別行為の道徳的責任を考えるさいには、先に見た事実的無知の構造がそのままあてはまることになる。つまり、その差別行為の責任があるとは、そのひとの道徳的無知（差別行為が不正であることに気づいていないこと）が責められるべき無知であるということにほかならない。そしてそれが責められるべき無知であるのは、その無知をもたらした浅はかな行為が非難に値する場合にかぎられるのである。

これが「無知からの行為の道徳的責任」の基本図式である。Rosen（2004）が強調するように、この図式によると、無知からなされた行為の責任そのものは、派生的な責任 derivative responsibility とみなされることになる（Rosen, 2004, p. 300）。というのも、無知からなされた行為についてそのひとに責任があるのは、その無知をもたらした浅はかな行為についてそのひとに責任がある場合にかぎられるからだ。したがって本来の責任 original responsibility の場所は、過去の浅はかな行為へと遡るのである。

3. 潜在的態度

前節で見た「無知からの行為の責任」をめぐる図式のなかで、潜在的態度という心的状態はどのように無知と類比的に位置づけられうるのだろうか。そのことを考察する前にまず、潜在的態度はどのように測定されて、どのような特徴を備えているのかを簡潔に見ておくことにしよう。

あるひとがある特定の人種に属する人々をどのように評価しているかを調べたいとしてみよう。この場合、質問紙に答えてもらうように、本人の自己報告に頼ることは有望なやりかたではない。たとえばそのひとは、差別は不正であるという一般に望ましいとされる考えをふまえて、それに沿った回答をしてしまうかもしれないからだ。しかし近年、そのような困難を克服するとされる実験手法が開発され、さまざまな成果を挙げている。本節ではまず、その代表的な手法である潜在連合テスト Implicit Association Test (IAT) (Greenwald, McGhee, & Schwartz, 1998) を紹介し、それによって測定される心的状態（潜在的態度）にどのような特徴があるかとされるのかを概観する。

IAT は、概念どうしの連合の強さを間接的に測定するために開発された手法であり、そのアイデアを説明するためにしばしば次のようなトランプのカードを分類する課題が引き合いに出される (e.g., Greenwald and Farnham, 2000, pp. 1022-1023)。52 枚のトランプのカードから「スペードかクラブ」と「ハートかダイヤ」をそれぞれ左側と右側に振り分ける課題と、「スペードかダイヤ」と「ハートかクラブ」をそれぞれ左側と右側に振り分ける課題を比べてみよう。実際にやってみるとわかるが、前者の課題のほうが後者よりも容易であるはずだ。というのも、スペードとクラブ、ダイヤとハートには、それぞれの色（赤色と黒色）によって強い連合がもたらされるからだ。このアイデアをもとに、たとえば人種とその評価の連合を調べたい場合、「白人に典型的な名前か快語」と「黒人に典型的な名前か不快語」を分類する課題と、「白人に典型的な名前か不快語」と「黒人に典型的な名前か快語」を分類する課題の、それぞれの容易さを比較することで、被験者が抱いている人種（黒人ないし白人）とその評価（ポジティブないしネガティブなバイレンス）の連合を調べることができる (e.g., Greenwald, McGhee, & Schwartz, 1998)。このようにして測定される心的状態が潜在的態度である。

IAT は、ジェンダー (e.g., Rudman & Kilianski, 2000) や若年者／高齢者 (e.g., Nosek, Banaji, & Greenwald, 2002) など、さまざまな集団にたいする潜在的態度を調べるために実施されてきた。ここでは人種 IAT を例として取り上げて、潜在的態度にそなわる特徴として次節の議論に密接にかかわるふたつの特徴に言及しておこう。

第一に、IAT によって測定される潜在的態度は、自己報告を通じてなされる回答と対立しうる。つまり、平等主義的な立場に関与していることが自己報告で明言されながらも、他方で特定の人種をネガティブに評価する潜在的態度が同時にともなっていることが IAT において示されることがある (Greenwald, et al., 2009)。

第二に、人種差別のように社会的にセンシティブなトピックについては、自己報告よりも IAT のほうが、そこから生じる行動の予測においてすぐれていることがわかっている (Greenwald, et al., 2009)。その範囲は、会話をいつ切り上げるかとか、相手の話にどれくらい笑うかといった、ふだんのちょっとしたやりとり (McConnell & Leibold, 2001) から、病気の治療法の選択といったシビアなもの (Green et al., 2007) まで、多岐にわたる。

4. 潜在的態度と道徳的責任

前節で紹介した潜在的態度の特徴をふまえると、次のような問題を考えることができるだろう。あるひとは、顕在的には平等主義的な信念を抱いていながら、同時に、特定の人種をネガティブに評価する潜在的態度を抱いていることがIATによってわかったとしてみよう。その潜在的態度の影響で差別行為がなされるとき、その行為について当人に道徳的責任はあるのだろうか。

以下では、このような「潜在的態度からの行為の責任」を考えるうえで、第2節で考察した「無知からの行為の責任」についての基本図式がどこまで適用できるのか（以下ではこれを「アナロジーの問題」と呼ぶことにしよう）を考える。そのさい、アナロジーの問題にとって重要なふたつの論点を明確にする。ひとつは潜在的態度と気づきの欠如の関係であり（4.1.）、もうひとつは潜在的態度と過去の浅はかな行為の関係（4.2）である。

4. 1. 気づきの欠如

道徳的責任の文脈において潜在的態度は無知とおなじように気づきの欠如として捉えることができるのか、という論点からはじめることにしよう。

前節で見たように、IATによって測定される潜在的態度と自己報告による顕在的態度とのあいだにはギャップが生じることがある。しかしそれはなぜだろうか。ひとつの自然な考えは、自己報告が意識的な心的状態を反映するのに対しIATは無意識の心的状態を反映する、というものだ。つまりそのようなギャップが生じるのは、その内容に（ふつうは）気づかれていない心的状態をIATが測定しているからだと説明されうるのである（e.g., Nosek, 2007, p. 68）。

他方で無知とは、気づきの対象となる特定の内容が欠如していることにほかならない。たとえば先述の無知な医師の場合、安全な代替薬があるという内容がそもそも存在しないのだから、当然その内容には気づかれていない。それゆえ、無知と潜在的態度の両方を、内容の気づきの欠如として捉えて、アナロジーの問題にとりくむためのひとつのステップとすることができるかもしれない。そのばあい第2節の図式は、特定の内容に気づいていないことがともなう行為の責任というかたちで一般化されることになるだろう。たとえば以下のSaul (2013)の主張を見てみよう。

被差別集団に対するバイアスはすべて非難に値するという見解を放棄することも重要だと思われる。私がこの見解を放棄する第一の理由は、それが偽であることにある。ひとは、ただ自分が性差別主義的な文化に住んでいるという事実だけの結果として生じる、自分が完全に気づいていない潜在バイアスについて非難されるべきではない（Saul, 2013, p. 55）

このようにSaulは、潜在的態度の内容に気づいているかどうかと、そこから生じる行為が非難に値するかどうかを結びつける²。そして第2節の図式に照らすと、この潜在的態度が責められるべきかどうかは、それをもたらした過去の浅はかな行為によって決まることになる。たとえば上記のSaulの主張を検討するためには、そこで言及されている「性差別主義的な文化に住んでいる」という事実が、非難に値する浅はかな行為となりうるかどうかを明確にすな

くてはならない。この点については4.2で論じることしよう。

しかし近年、潜在的態度にはその内容の気づきがともなうという証拠が蓄積されつつある。たとえば Hahn et al. (2014) では、被験者に自身の IAT のパフォーマンスを予測するよう事前に求めたうえで IAT を受けさせると、その予測がかなり正確であることが示された。このことは、被験者が潜在バイアスの内容に気づいていたことを強く示唆する。

それにたいしては、道徳的責任の文脈において無知が重要になるのは、内容に気づいていないこととは異なる意味での気づきの欠如だと答えることもできる。つまり、無知が道徳的責任の問題をもたらすのは、ある特定の内容に気づいていないからではなく、むしろその内容にもとづいて意思決定ができないからだという応答も可能である。たとえば先述の医師の場合、(無知のせいで) 代替薬を用いることが正しいという決断が下すことができなかった。そしてまさにそのために、医師の行為の責任が問題になったと考えるのはもっともらしい。このことをふまえると、私たちが潜在的態度の内容に気づいているとしても、その内容が意思決定や行動にどう影響するのかに気づいていないのだとすれば、依然としてアナロジーの問題にとりくむ余地は残されるだろう。たとえば、その顕在的信念のために相手を平等に扱う強い動機がありながら、その信念と対立する潜在的態度に影響されて差別行為をしてしまうひとを考えてみよう。いまの議論をあてはめると、そのひとは自分の潜在的態度の内容に気づいていたし、自分の行為をコントロールすることもできたが、その潜在的態度が意思決定(およびその後の行為)にどう影響するかに気づいていなかったために差別行為をおこなってしまった、と解釈することができる。

Gawronski, Hofmann, and Wilbur (2006) はこのような意味での気づき(の欠如)を、先に見た「内容の気づき content awareness」から区別して「インパクトの気づき impact awareness」と呼ぶ³。そして経験的研究に言及することで、私たちはじっさい潜在的態度の内容に気づいているが、他方でそのインパクトには気づいていないと結論する。これが正しいとすれば、無知からの行為と潜在的態度からの行為はどちらも、そのインパクトに気づいていないことがともなう行為として一般化することができるだろう。

Levy (2014 a, 2014 b) が特徴づける潜在的態度は、このインパクトの気づきの欠如という観点から捉えることができるように思われる⁴。Levy によると、顕在的態度と潜在的態度の違いは、その内容に気づいているかどうかにあるというよりむしろ、さまざまな反応をする傾向性とその内容のつながりかたにあるとされる。つまり、一方で信念のような顕在的態度とつながる傾向性は、推論の規範に照らして私たちがもつべき傾向性とおなじ構造をしているが、他方で潜在的態度と傾向性のつながりは命題的というより連合的で雑多なものであり、推論を通じた意思決定にともなうようなものではないとされるのである(e.g., Levy, 2014 b, pp. 31-32)。このような違いは、問題の態度が自身の判断にどう影響するかに気づいているかどうか(つまりそのインパクトに気づいているかどうか)によって説明されうるかもしれない。すなわち潜在的態度の場合、そのインパクトに気づいていないために、さまざまな傾向性と緩やかにしかつながることができないのである。

他方 Holroyd (2015) は、Monteith, Voils, and Ashburn-Nardo, (2001)の研究に依拠することで、私たちがときに自分の潜在的態度のインパクトに実際に気づいていることがあると論じる(Holroyd, 2015, pp. 518-520)。この主張が正しいとすれば、潜在的態度は全体としてはインパクトの気づきの欠如という点で無知と類比できるわけではなく、むしろ、潜在的態度の

なかに無知に含まれるものがあるとみなされることになる⁵。そしてこのことは潜在的態度そのものを無知と類比的に捉えようとする動機を損なうことになるだろう。

4. 2. 浅はかな行為

アナロジーの問題にはもうひとつの重要な論点がある。無知が責められるべきものであり、そこから生じる行為に責任を負わされるのは、第2節の図式にしたがうと、非難に値する浅はかな行為が過去に存在したどうかに帰着することを思い出そう。潜在的態度からの行為にもおなじことがあてはまるのだろうか。

ここで Rosen が無知からの行為の責任を論じるさいに、先に見た Saul とおなじように性差別主義者の事例に言及しているのは興味深い（Rosen, 2003, pp. 66-67）。自分の息子と娘を平等に扱わないがその行為が不正だとは気づいていない、1950年代の性差別主義者のアメリカ人の父親をとりあげて、Rosen はこの父親の行為が（不正ではあるが）非難に値しないと論じる。その理由はこうだ。問題の父親には、自分のおこないが許されるかどうかにかかわる情報を探求する「手続き上の認識的義務 procedural epistemic obligation」（Rosen, 2004, p. 301）がある。Rosen によれば、この義務があることをわかっていながらそれに違反し、その結果として無知が生じるとき、当人は非難に値する浅はかな行為をしたことになる⁶。しかし1950年代のアメリカ人の父親が、探求を通じて無知を防ぎ、自分の行為が不正であると気づくようになるためには、当時の事情をふまえると「途方もない道徳的な洞察」（Rosen, 2003, p. 69）が必要であり、それはあまりに厳しい要求だろう。それゆえ、無知を責められるべきものにするような、非難に値する浅はかな行為が過去に存在したわけではなく、結果として、その無知から生じる不正な行為（自分の娘と息子を差別して扱うこと）もまた、非難には値しないとされるのである。

「ただ自分が性差別主義的な文化に住んでいるという事実だけの結果として生じる」潜在的態度は非難に値しないという、先に引用した Saul の主張は、この Rosen の議論にほぼ対応したものと述べることができるだろう。Saul が想定するように、私たちが潜在的態度の内容に気づくことができないのだとすれば（あるいは気づいていたとしても、そのインパクトに気づくことができないとすれば）、その気づきの欠如を防ぐための認識的義務を課することはきわめて厳しい要求だと考えられるだろう。さらに、もし潜在的態度と対立する平等主義的な顕在的態度が自身の熟慮によってもたらされているのだとすれば、ある意味では、ふつうに考えられる認識的義務を満たしていると述べることもさえてできるかもしれない。

他方で Washington and Kelly (2016) は、潜在的態度の内容に内省を通じてアクセスできないと想定するが、Saul とは異なり、それでも潜在的態度から生じる行為の責任を負わせることのできるケースがあると論じる。その主張のポイントは、潜在的態度の内容に気づく道筋は内省にかぎらないという点にある。実際いまの私たちは（たとえば1950年代のアメリカ人の父親と異なり）IAT といった「外部の道筋」（Washington & Kelly, 2016, p. 25）を通じて、自分の潜在的態度の内容に間接的に気づくことができる。それゆえ行為者の社会的役割（たとえば親であるとか医療関係者であるとか企業の雇用担当者であるといった役割）によっては、そのような態度を防ぐための認識的義務を事前に果たすべきだったと論じられる⁷。つまり、おなじ潜在的態度であっても、それを防ぐために課される認識的義務は、行為者の社会的役割やその時代の認識環境といった、行為者の外部の文脈によって変わりうるとされるのである。

無知からの行為を論じるさいの Rosen の主張によると、そういった行為の責任は過去の浅はかな行為が非難に値する場合に生じ、その浅はかな行為が非難に値するのは、問題の無知（およびそこから生じる行為）を防ぐための認識的義務をわかったうえで果たさなかったときだとされる。Saul と Washington and Kelly の議論はいずれも（その結論は異なるとはいえ）それと類比的なかたちで、潜在的態度からの行為を捉えようとするものだともみなすことができる。それゆえこれらは、アナロジーの問題に肯定的に答えようとする立場として理解できるだろう。

5. おわりに

無知からの行為と潜在的態度からの行為は、その道徳的責任を考えるうえで、類比的に捉えることができるのだろうか。このアナロジーの問題に肯定的に答えることができるかどうかには、ふたつの論点がともなう。ひとつは「潜在的態度は気づきの欠如とみなすことができるのか」という論点であり、もうひとつは「潜在的態度からの行為の責任は過去の浅はかな行為に由来するのか」という論点である。これまで見てきたように、そのいずれの論点においても依然として議論の余地が残されている。

最後に今後の課題を述べておきたい。第2節で概観した図式は「通説 the Orthodoxy」(Wieland, 2017, p. 9) と呼ばれるように、無知からの行為の責任をめぐる議論の中核を形成する。それによると、無知からの行為の本来の責任の場所は、無知であること（あるいはそこから生じる行為）そのものにあるのではなく、それより過去の浅はかな行為へと遡るとされる。しかし無知からの行為の責任を考えるうえで、そのような通説を否定する立場もまた存在する。したがってアナロジーの問題にとりくむにあたっては、潜在的態度の責任を、そうした「通説」を否定する立場と類比的に理解しうるかどうかにについても考えなくてはならないだろう。たとえば、潜在的態度と無知とのあいだには気づきの欠如という点でアナロジーが成立するが、その責任は過去の浅はかな行為に由来するという（「通説」が主張する）点ではアナロジーが成立しない、という立場も可能かもしれない。

ここでは、そうした立場の事例に言及するだけに留めておく。ひとつは4.1 で見た Levy の主張である。それによると、潜在的態度はある種の気づきの欠如（インパクトの気づきの欠如）という点で無知と類比的なものだと解釈できた。しかし Levy はそのような潜在的態度からの行為の責任を論じるさい、過去の浅はかな行為に遡らずに、潜在的態度の構造そのものに焦点をあてる。つまり顕在的態度は、さまざまな傾向性と合理的に結びつくおかげで、道徳的行為者であることに必要な行動の統合性や一貫性をもたらしうのに対し、潜在的態度はその内容と傾向性のつながりがより緩やかであるために、そうした道徳的な行為者性をもたらさないといわれるのである (e.g., Levy, 2014 b, pp. 35-36)。それゆえ Levy の立場は（インパクトの）気づきの欠如という点で無知と潜在的態度を類比的に捉えながら、その責任を論じるさいに通説から逸脱することになる。また Holroyd (2015) は、4.1 で見たように、気づきの欠如という点で潜在的態度と無知とのあいだで全体的な類比は成立しないと考えているように思われるが、それに加えて、気づきが欠如している潜在的態度の責任を論じるさいにもまた、「通説」に明確に反対する立場 (e.g., Sher, 2009) に依拠している。これらの立場を分析することは今後の課題としたい。

注

- 1 本文の強調箇所には傍点を付した。引用については、原文がイタリック体ないし太字になっている箇所に傍点を付した。
- 2 Saul の主張は、潜在的態度そのものが非難に値するかどうかにかかわるのであって、潜在的態度から生じる行為に焦点をあてているわけではない（同様に潜在的態度そのものに焦点をあてる議論としては Kelly & Roedder (2008)を参照）。しかし第2節で見た基本的な図式のもとでは、無知やその原因となった過去の浅はかな行為が非難に値する程度は、その無知から生じる行為が非難に値する程度へと派生的に移行するとされる（Wieland, 2017, p. 10）。それゆえ、潜在的態度の場合にもおなじことが成り立つと考えるならば、Saul の主張はそのまま潜在的態度から生じる行為にも適用できるだろう。
- 3 Gawronski, Hofmann, and Wilbur (2006)はさらに「起源の気づき source awareness」という区分にも言及するが、ここでは取り上げない。
- 4 以下で述べる顕在的態度と潜在的態度の違いはコントロールの観点から説明されるかもしれないが、その場合でも、そうしたコントロールの違いはインパクトの気づきに依存していると述べることができるだろう。（e.g. Levy, 2017, p. 9）
- 5 実際 Holroyd (forthcoming)は、潜在的態度に気づいていないことが無知の一例にほかならないとみなしている。
- 6 無知からの行為の責任を論じるさいの重要な争点は、浅はかな行為がいつ非難に値するかということにある（Wieland, 2017, pp. 11-12）。Rosen の場合、認識的義務がわかったうえでそれを果たさないときだとされるのだから、浅はかな行為が非難に値するのは真正のアクラシアのケースにかぎられることになる（Rosen, 2004, p. 307）。しかし、浅はかな行為が非難に値するのはアクラシアのときにかぎられるかどうかには議論の余地がある。たとえば FitzPatrick (2008) は、その行為が「過信、傲慢、却下、怠惰、独断、無関心、わがまま、軽視といった悪徳を自発的に行使した結果」（FitzPatrick, 2008, p. 605）である場合もまた、それは非難に値すると論じる。
 このように、無知をもたらす浅はかな行為がいつ非難に値するかには議論の余地がある。それゆえアナロジーの問題にとりくむのであれば、この争点に応じたかたちで潜在的態度からの行為の責任を論じるべきだろう。このことは今後の課題として第5節で言及する。さしあたり本稿では紙幅の関係上、Rosen の提案する条件のみに焦点をあてることにする。
- 7 この認識的義務を果たすためには、たんに潜在的態度の存在に（間接的に）気づくだけでなく、それを適切にコントロールすることが求められるだろう。私たちが潜在的態度をどうコントロールしうるかについては、たとえば Holroyd (2012)がいくつかの経験的研究を検討したうえで論じている。

参考文献

- FitzPatrick, W. J. (2008). Moral responsibility and normative ignorance: Answering a new skeptical challenge. *Ethics*, 118(4), 589-613.
- Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are “implicit” attitudes unconscious? *Consciousness and Cognition*, 15(3), 485-499.
- Green, A., Carney, D., Pallin, D., Ngo, L., Raymond, K., Lezzoni, L., & Banaji, M. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, 22, 1231-1238.
- Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- & Farnham, S. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of personality and social psychology*, 79(6), 1022-1038.
- , Poehlman, T., Uhlmann, E., & Banaji, M. (2009). Understanding and using the Implicit Association Test:

- III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17-41.
- Hahn, A., Judd, C., Hirsh, H., Blair, I. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*, 143(3), 1369-1392.
- Holroyd, J. (2012). Responsibility for implicit bias. *Journal of Social Philosophy*, 43(3), 274-306.
- (2015). Implicit bias, awareness and imperfect cognitions. *Consciousness and Cognition*, 33, 511-523.
- (forthcoming). Two ways of socialising responsibility: Circumstantialist and scaffolded-responsiveness. In K. Hutchinson, C. MacKenzie, & M. Oshana (Eds.), *Social Dimensions of Moral Responsibility*. Oxford University Press.
- Kelly, D., & Roedder, E. (2008). Racial cognition and the ethics of implicit bias. *Philosophy Compass*, 3(3), 522-540.
- Levy, N. (2014 a). *Consciousness and Moral Responsibility*. Oxford University Press.
- (2014 b). Consciousness, implicit attitudes and moral responsibility. *Noûs*, 48(1), 21-40.
- (2017). Implicit bias and moral responsibility: Probing the data. *Philosophy and Phenomenological Research*, 94(1), 3-26.
- McConnell, A., & Leibold, J. (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37, 435-442.
- Monteith, M., Voils, C., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting and reacting to implicit racial biases. *Social Cognition*, 19(4), 395-417.
- Nosek, B. (2007). Implicit-explicit relationships. *Current Directions in Psychological Sciences*, 16, 65-69.
- , Banaji, M., & Greenwald, A. (2002). Harvesting implicit group attitudes and beliefs from demonstration website. *Group Dynamics*, 6(1), 101-115.
- Rosen, G. (2003). Culpability and ignorance. *Proceedings of the Aristotelian Society*, 103(1), 61-84.
- (2004). Skepticism about moral responsibility. *Philosophical Perspectives*, 18(1), 295-313.
- Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin*, 26(11), 1315-1328.
- Saul, J. (2013). Implicit bias, stereotype threat, and women in philosophy. In K. Hutchinson & F. Jenkins (Eds.), *Women in Philosophy: What Needs to Change?* (pp. 39-60). Oxford University Press.
- Sher, G. (2009). *Who Knew? Responsibility Without Awareness*. Oxford University Press.
- Smith, H. (1983). Culpable ignorance. *The Philosophical Review*, 92(4), 543-571.
- Washington, N., & Kelly, D. (2016). Who's responsible for this? In M. Brownstein & J. Saul (Eds.), *Implicit Bias and Philosophy vol.2* (pp. 11-36). Oxford University Press.
- Wieland, J. (2017). Introduction. In P. Robichaud & J. Wieland (Eds.), *Responsibility: The Epistemic Condition* (pp. 1-28). Oxford University Press.