Francesco Chiesa

Responsabilità politica per il bias implicito*

1. Introduzione

Con "bias implicito" s'intendono quei casi in cui uno stereotipo su un gruppo x viene proiettato da un agente y automaticamente a singoli membri del gruppo x influenzando il comportamento di y nei confronti di membri di x senza che y ne sia consapevole. La mia attenzione si concentra sui comportamenti influenzati dal bias implicito e sul tipo di responsabilità che possiamo (se possiamo) attribuire agli agenti per tali comportamenti quando hanno effetti discriminatori. Pertanto userò l'espressione "bias implicito" riferita alla sua manifestazione comportamentale e non al mero fatto che l'agente ospiti uno stato mentale *biased*. Non entro quindi nel dibattito sulla possibilità o meno di attribuire responsabilità morale agli agenti per le loro credenze[1] e stati mentali (Hieronymi 2008; Arpaly 2003). Inoltre, mi interessano i comportamenti involontariamente discriminatori indotti dal bias implicito in "interazioni pubblicamente rilevanti"; interazioni tra rappresentanti

* Una prima versione di questo articolo è stata presentata al seminario permanente sulla normatività pratica a Pavia il 9 aprile 2016. Ringrazio tutti i partecipanti per le domande e la discussione. In particolare, vorrei ringraziare Emanuela Ceva e Elisabetta Galeotti per aver letto e commentato con generosità il pezzo. Grazie anche a Federica Liveriero per il lavoro editoriale e per le chiacchierate su questo e altro.

[1] In questo lavoro utilizzo la parola "credenza" in un senso molto ampio che include stati mentali cognitivi e affettivi più o meno strutturati. Avere bias implicito non significa credere P in senso proposizionale (le credenze proposizionali sono predicati di verità e falsità come "la neve è bianca".) Il bias implicito prende forma a un livello cognitivo e affettivo più profondo, spesso inconscio.

di istituzioni pubbliche e cittadini e tra concittadini nella sfera pubblica. Tralascio invece di indagare l'effetto del bias implicito nelle relazioni private, per esempio nella scelta delle persone con cui uscire a cena, né se queste scelte abbiano rilevanza morale.

In questo articolo sostengo che l'agente che si comporta in modo discriminatorio a causa del bias implicito non può essere considerato moralmente responsabile nel senso noto in letteratura come *attributability*, e neppure in alcuni sensi di *accountability*[2]. Si ha attributability quando possiamo collegare un'attitudine o azione all'identità pratica di chi la compie e ai suoi impegni valoriali (Watson 1996). L'agente è responsabile nel senso di *attributability* se possiamo dire che in un grado rilevante le sue azioni svelano il suo sé reale e profondo; che le sue azioni sono, in grado rilevante, in sintonia con la sua generale prospettiva valoriale (Frankfurt 1971; Scanlon 1998, 277-294; Levy 2011; Smith 2005, 2008; Glasgow 2016). Nel paragrafo 4 sostengo che il bias implicito non soddisfa le condizioni standard della responsabilità morale come *attributability*. Il bias implicito suscita controversia morale precisamente perché non sembra rispecchiare il "sé reale" e può rappresentare un caso di «disarmonia interna» (Gendler 2008b) tra le credenze esplicite e quelle implicite. Il concetto di *accountability* riguarda invece la possibilità di imputare le conseguenze di un'azione a un agente in senso causale e non agenziale. L'attenzione in questo caso non è sulla prospettiva valoriale dell'agente, ma sulle conseguenze delle sue azioni, indipendentemente da considerazioni di lode e biasimo. Essa è solitamente invocata per allocare oneri e benefici tra diversi ruoli sociali nella comunità politica.

La mia cauta conclusione sull'attribuzione di responsabilità morale per bias implicito non deve però spingerci tra le braccia inattive dello scetticismo morale, considerando così il comportamento discriminatorio del bias implicito come una calamità naturale rispetto alla quale nulla possiamo. Esiste infatti una concezione di *accountability* che può aiutarci a capire cosa c'è di ingiusto nel bias implicito. Sappiamo che il bias implicito moralmente rilevante si nutre di condizioni strutturali esterne all'agente come standard sociali, norme sociali e stereotipi prevalenti nel contesto sociale in cui vive l'agente. Ciò che rende il bias implicito ingiusto è il fatto che sia causato da strutture sociali ingiuste. Sostengo quindi che le istituzioni politiche hanno la responsabilità prospettica (*forward-looking*) di incidere su quegli standard, norme sociali e stereotipi che nutrono il bias implicito. I cittadini hanno la responsabilità indiretta di farsi sentire come membri della comunità politica

[2] Al fine di evitare inutili parafrasi e di non appesantire il testo con traduzioni ambigue, userò questi due termini prevalentemente usati nella letteratura di riferimento.

affinché le istituzioni soddisfino tale dovere. Intendo distinguere il bias implicito "moralmente rilevante" dal bias implicito meramente individuale e idiosincratico con profilo eziologico tutto psicologico e non socio-politico.

La tesi è, quindi, che non possiamo attribuire responsabilità diretta a individui specifici, ma possiamo attribuire alle istituzioni la responsabilità di trasformare l'ambiente epistemico in cui il bias implicito si forma e si nutre. A supporto di questo argomento, propongo una rivisitazione della condizione epistemica della responsabilità come esterna all'individuo e come una condizione epistemica sociale diffusa che le istituzioni possono e devono influenzare.

2. Cosa sappiamo sul bias implicito

La ricerca empirica sul bias implicito è ormai vastissima. Sappiamo, in modo abbastanza condiviso, che la maggior parte di noi, anche gli egualitari più convinti e sinceri, può essere vittima di bias implicito nei confronti di appartenenti a gruppi che differiscono dalle norme sociali prevalenti (Huebner 2016). Sappiamo anche che il bias implicito viene tendenzialmente acquisito dalle strutture sociali in cui viviamo[3], dagli stereotipi culturali, dalle ideologie, dai modelli di vantaggio e di svantaggio che ci stanno intorno (Saul 2013a; Banaji e Greenwald 2013).

Jennifer Saul descrive il bias implicito come «unconscious biases that affect the way we perceive, evaluate, or interact with people from the groups that our biases target» (2013b, 40). Jules Holroyd e Joseph Sweetman scrivono che «implicit bias are whatever unconscious processes influence our perceptions, judgements and actions – in this context, in relation to social category members (women, blacks, gays, for example)» (2016, 81). Per Keith Frankish, le persone sono affette da bias implicito se

> their behaviour manifests a stereotyped conception […] even if they do not explicitly endorse the conception and perhaps explicitly reject it. The possibility of such implicit bias is a matter of ethical concern, since it means that bias may persist in an unacknowledged, 'underground' form, even when it has been explicitly repudiated (Frankish 2016, 24).

Carole J. Lee descrive il bias implicito come «triggered automatically even in the absence of explicit attention being drawn to the stereotype-relevance of per-

[3] Con strutture sociali intendo vari fenomeni sociali come le istituzioni, le convenzioni, le pratiche e le norme sociali, i ruoli sociali e le ideologie (Haslanger 2012, 413-418).

ceptual cues across a range of different perceptual modalities» (2016, 273). Holroyd definisce così il fenomeno:

> An individual harbors an implicit bias against some stigmatized group (G), when she has automatic cognitive or affective associations between (her concept of) G and some negative property (P) or stereotypic trait (T), which are accessible and can be operative in influencing judgement and behavior without the conscious awareness of the agent (2012, 275).

Con «accessible» Holroyd non intende accessibile all'introspezione, ma accessibile nei termini della velocità e della facilità con cui l'associazione automatica viene fatta (Holroyd 2012, 302, n. 2).

La ricerca empirica ci mostra che il bias influenza i comportamenti in una varietà preoccupante di circostanze. Per esempio, lo stesso *curriculum vitae* ottiene una diversa considerazione se è associato a un nome tipicamente bianco anziché a uno tipicamente nero, o a un nome maschile anziché a un nome femminile (Dovidio e Gaertner 2000; Bertrand e Mullainathan 2004; Krieger 1995; Steinpreis *et al.* 1999). Studi di laboratorio sul cosiddetto *shooter bias* mostrano come il bias implicito possa influenzare la percezione: lo stesso oggetto dall'apparenza ambigua ha più probabilità di essere visto come una pistola se è in mano a un uomo nero e come qualcos'altro (spesso un telefono) se è in mano a un uomo bianco (stessi effetti sulla percezione sono stati dimostrati su uomini di apparente religione islamica e altri che apparivano come non islamici). I partecipanti a questi studi tendono con maggior probabilità a non sparare all'uomo bianco armato rispetto all'uomo nero armato (Unkelbach *et al.* 2008; Correll *et al.* 2006; Correll *et al.* 2002; Payne 2001). Ci sono studi su come il bias implicito può influenzare azioni di controllo di polizia di cosiddetto *stop and frisk* che operano secondo la logica del «sospetto ragionevole» (Clemons 2014). Studi sull'impatto del colore della pelle su come i giudici e i giurati nei tribunali americani valutano l'evidenza (Levinson e Young 2010; Levinson 2007; Rachlinski *et al.* 2009). E poi ancora studi su come il bias implicito può influenzare la valutazione degli studenti (Bradley 1993); studi su come il bias sul prestigio dell'affiliazione istituzionale può influenzare le *submission* a riviste accademiche (Peters e Ceci 1982); studi su come gli studenti tendano a valutare in modo diverso lo stesso insegnate virtuale di un corso online che a un campione si presenta con identità maschile e a un altro con identità femminile (MacNell *et al.* 2015); studi su come cambia la percentuale di donne nelle orchestre sinfoniche quando le audizioni vengono fatte dietro schermi opachi (Goldin e Rouse 1997). Ancora, la ricerca empirica sul cosiddetto *sequential priming*, che misura l'attivazione automatica del

bias e, per esempio, espone soggetti a etichette di gruppo come "black" o "woman" e misura i loro tempi di risposta (*response latencies*) a parole stereotipiche come "lazy" e "nurturing" (Fazio 1995). Ci sono anche studi su come il personale medico possa trattare i pazienti in modo inconsapevolmente diverso in base al colore della pelle (White e Chanoff 2011). Per esempio, un recente studio ha mostrato come i bambini neri con appendicite tendono a ricevere meno antidolorifici dei bambini bianchi con la stessa patologia, anche quando i medici percepiscono che il livello di dolore sia analogo (Goyal *et al.* 2015). Inoltre, c'è la tendenza a raccomandare meno approfondimenti medici a pazienti neri che ai pazienti bianchi, anche se in possesso degli stessa cartella clinica (Schulman *et al.* 1999).

Ancora, un recente studio randomizzato *double blind* ha mostrato che

> science faculty from research-intensive universities rated the application materials of a student – who was randomly assigned either a male or female name – for a laboratory manager position. Faculty participants rated the male applicant as significantly more competent and hireable than the (identical) female applicant. These participants also selected a higher starting salary and offered more career mentoring to the male applicant. The gender of the faculty participants did not affect responses; both female and male faculty members were equally likely to exhibit bias against the female student (Moss-Racusin *et al.* 2012, 16474).

Il fatto che i selezionatori uomini e donne avessero la stessa propensione a sfavorire la medesima candidatura quando presentata sotto identità femminile indica un tema ricorrente nella letteratura sotto esame: il bias implicito non è un caso (o non è solo un caso) di favoritismo per il proprio gruppo (*in-group favouritism*) ma si tratta perlopiù di tendenze sociali e stereotipi che si acquisiscono nel contesto sociale in cui si vive che si riproducono in modo inconscio nel proprio comportamento. Quindi, il bias implicito che rispecchia stereotipi esistenti su donne e afroamericani può essere acquisito e riprodotto inconsciamente da donne e afroamericani. L'implicazione politica nel caso del nostro esempio del *science faculty committee* è che questi casi di bias implicito non possono essere semplicemente affrontati con quella che Anne Phillips chiama la «politica della presenza» (1998) (che in questo caso significherebbe aumentare il numero di donne nella commissione di selezione). Maggior rappresentazione dei gruppi e delle identità, per quanto possa essere una pratica virtuosa per altre ragioni, non riduce necessariamente questo tipo di discriminazione (si veda Milkman *et al.* 2015).

2.1. Bias implicito e attitudini esplicite

Nonostante l'evidenza empirica[4] dimostri che il comportamento discriminatorio accade, in una certa misura, sotto il livello di coscienza e fuori dal nostro radar introspettivo (e, quindi, implicitamente), non tutti questi comportamenti discriminatori sono necessariamente solo prodotti dal bias implicito. Alcuni possono anche avere un bias esplicito; in questi casi diremo che il bias implicito non è contrario alle credenze esplicite, mentre è avverso il bias che non verrebbe approvato qualora venissimo a conoscenza di esso. Poiché, come specificherò nel prossimo paragrafo, il caso che intendo discutere è quello dell'egualitario con bias implicito, il tipo di bias implicito che mi interessa è proprio quello avverso.

A differenza degli atteggiamenti espliciti, che sappiamo di avere e siamo in grado di verbalizzare, il bias implicito può modificare il comportamento senza che ne siamo consapevoli, e senza che tale influenza sia potenziale oggetto di nostra introspezione. Gli atteggiamenti espliciti e quelli impliciti possono essere diretti verso lo stesso oggetto ed essere discordanti tra loro. Posso, per esempio, avere credenze genuinamente egualitarie e antirazziste e comportarmi in modo discriminatorio con persone dal colore della pelle di un certo tipo senza rendermene conto (Fazio e Olson 2003; Stanley *et al.* 2008; Gawronski e Payne 2010; Nosek *et al.* 2011).

È stato dimostrato che il bias implicito prevale sugli atteggiamenti espliciti nel determinare comportamenti e giudizi in contesti sociali rilevanti (Faucher 2016, 116; Greenwald *et al.* 2009; Pearsons *et al.* 2009). Secondo alcuni, il bias implicito si forma in modo completamente indipendente dalle convinzioni esplicite (Saul 2013a; Holroyd e Sweetman 2016, 89); secondo altri (Devine *et al.* 2002) ci sarebbe qualche correlazione tra il grado della manifestazione comportamentale del bias implicito e le credenze esplicite dell'agente (Nosek *et al.* 2005). Probabilmente, entrambe le cose possono essere vere in casi diversi, data la natura varia del

[4] Tra gli strumenti che tentano di misurare le associazioni implicite degli individui il più famoso è forse l'Implicit Association Test (IAT). Per una rassegna del dibattito metodologico si vedano Jost *et al.* (2009) e Greenwald *et al.* (1998). Sulla validità predittiva dello IAT rispetto agli effettivi comportamenti discriminatori, si vedano Greenwald *et al.* (2003); Greenwald *et al.* (2009); Nosek *et al.* (2005); Lane *et al.* (2007); Oswald *et al.* (2015). Un altro test di misurazione è l'*affective priming test* (Fazio *et al.* 1995) che rivela la velocità con cui associamo concetti e categorie in modo che non espliciteremmo, sia per ragioni di desiderabilità sociale (Fazio e Olson 2003) (e questo non è un caso di bias implicito), oppure perché (nel caso di bias implicito) tali associazioni non sono rilevabili dall'agente (Holroyd e Sweetman 2016, 85; Brownstein e Saul 2016a).

fenomeno che Holroyd e Sweetman definiscono come "funzionalmente etereogeneo» (2016).

La distinzione implicito-esplicito poggia su un paradigma dominante in psicologia sociale e nelle scienze cognitive che spesso viene chiamato "processo duale" o "modello duale di sistema". L'idea è che la nostra mente sia formata da due flussi di elaborazione delle informazioni; uno lento, controllato, conscio, deliberativo e inferenziale (fonte del pensiero astratto e ipotetico); il secondo è rapido, automatico, associativo, inconscio e affettivo (Chaiken e Trope 1999; Gendler 2008b; Bargh e Chartrand 1999; Kahneman 2011).

Quindi, il bias implicito viene tipicamente considerato come prodotto al di fuori del nostro controllo e consapevolezza. Sebbene sia pacifico che il bias implicito sia cosa diversa da credenze esplicite consapevolmente detenute, c'è un dibattito sulla questione se le associazioni implicite siano semantiche o affettive, o entrambe, e su cosa esattamente si intende quando si dice che il bias implicito è inconscio (Holroyd e Sweetman 2016; Machery 2016). Alcuni identificano la componente implicita con l'automaticità (per esempio Fazio *et al.* 1995; Devine 1989) e distinguono tra modelli di elaborazione dell'informazione nella memoria "controllati" da quelli "automatici" (Shiffrin e Schneider 1977; Huebner 2016, 49). Un secondo filone di ricerca identifica la componente implicita con la dimensione inconscia (per esempio Banaji e Greenwald 2013). In questo caso, l'enfasi non è posta sull'automaticità ma sull'indisponibilità introspettiva del bias implicito (Brownstein e Saul 2016a, 8).

Nella letteratura sul bias implicito, le categorie di "inconscio" e "inconsapevole" sembrano spesso usate in modo interscambiabile, per quanto, anche a un livello colloquiale e intuitivo, tendiamo a pensare alle due cose come non necessariamente identiche. La letteratura tende a favorire la posizione secondo cui, se diventassimo consapevoli del nostro bias implicito, ciò avrebbe effetto a livello cognitivo, ma non necessariamente comportamentale: quand'anche sapessi con una certa probabilità di avere il bias implicito come stato mentale potrei, allo stesso tempo, essere completamente inconsapevole (anche nel senso di inconscio) di quando e come tale contenuto cognitivo e affettivo impatti sul comportamento. C'è infatti evidenza empirica del fatto che aumentare la consapevolezza della probabilità di detenere un bias implicito non si traduce in una maggiore abilità di comportarsi in modo meno discriminatorio (Gawronski *et al.* 2006).

In questo articolo tendo a riferirmi alle nozioni di inconsapevolezza e incoscienza del bias implicito in modo un po' sfocato perché tale è lo stato dell'arte della letteratura sul tema.

3. Se sei egualitario, come mai sei così *biased*?

Come ho detto, secondo un numero crescente di studi, il bias implicito si può tradurre in comportamenti effettivi, anche se uno è genuinamente convinto e impegnato a rispettare principi egualitari, specialmente in particolari circostanze (Pearson *et al.* 2009).

Concentro la mia analisi sull'egualitario autentico che si comporta in modo involontariamente discriminatorio. Il mio tipo egualitario è un "razzista o sessista avverso", vale a dire qualcuno con nessun pregiudizio razzista esplicito e tuttavia con alto bias implicito: un agente che, qualora ne diventasse consapevole, disconoscerebbe il proprio bias. Al contrario del razzista avverso, il razzista esplicito solitamente unisce alle credenze razziste esplicite il bias implicito. È pienamente consapevole delle sue credenze e non ne è disturbato; anzi, le avvalla in modo riflessivo e agisce sulla loro base quando sorge l'occasione. Il mio tipo, invece, è un egualitario esplicitamente antirazzista (ossia avvalla sinceramente il proprio antirazzismo se ci riflette su), ma ospita del bias implicito razzista con implicazioni comportamentali.

Si considerino questi tre diversi personaggi (il razzismo qui è usato come esempio, sostituibile con sessismo, classismo e così via):

Donald ha credenze esplicite razziste e bias implicito razzista; non ne è disturbato e discrimina, quando è il caso.

Giorgia ha credenze esplicite razziste e bias implicito razzista; vorrebbe però non discriminare, ma discrimina.

John ha credenze esplicite egualitarie, ma un bias implicito razzista; non vuole discriminare, ma involontariamente discrimina.

John è il protagonista di questo articolo. La domanda che pongo è se John stia facendo qualcosa di moralmente biasimevole. Dovremmo attribuire responsabilità morale (e se sì, in che senso) a John per il suo comportamento discriminatorio involontario? John sta commettendo un'ingiustizia e, se sì, di che tipo? Prima di tentare una risposta a queste domande, due parole sul perché non mi occupo dei casi di Donald e di Giorgia.

Donald è un caso standard di razzista volontario che avvalla consapevolmente le sue (esplicite) credenze razziste, e qualora diventasse consapevole di avere anche un bias implicito razzista, rivendicherebbe la coerenza interna tra il suo mondo conscio e quello inconscio. Anche se può intervenire per frenare e reprimere il proprio razzismo esplicito, minimizzando così il proprio comportamento discriminatorio, non ha alcuna intenzione di farlo. Donald nutre il proprio razzismo esplicito soddisfacendo pertanto le condizioni standard per l'attribuzione della responsabilità morale – conoscenza e volontarietà – di cui dirò tra poco.

Giorgia è un caso moralmente più interessante perché discrimina in modo involontario. Tuttavia, la fonte della sua discriminazione è sia esplicita sia implicita. È una razzista consapevole delle proprie credenze razziste, ma in certi contesti Giorgia desidera non discriminare per nascondere il proprio razzismo per ragioni di desiderabilità sociale. Diciamo che è convinta di dover fare buon viso al cattivo gioco di ciò che considera essere l'ipocrita correttezza politica dei suoi colleghi per non essere stigmatizzata come razzista al lavoro. Tuttavia, il suo tentativo fallisce: in un momento di stress o di relax, la vera Giorgia prende involontariamente il sopravvento e il suo bias implicito emerge in un comportamento discriminatorio. Giorgia, come Donald, avvalla consapevolmente il proprio razzismo ma, a differenza di Donald, è ambivalente sui comportamenti discriminatori. Per ragioni strategiche, come la necessità di conformarsi alla pressione dei pari al lavoro, cerca di agire come se non fosse razzista. Il suo comportamento discriminatorio è involontario solo in senso contingente; Giorgia non si sente alienata rispetto ai suoi convincimenti comportandosi in modo razzista; si dispiace di averlo fatto di fronte ai colleghi durante una riunione per selezionare nuovo personale perché non vuole render nota pubblicamente questa sua posizione. La sua intenzione di non discriminare non era motivata da credenze egualitarie che non ha, ma da preoccupazioni di autopresentazione sociale. Nascondere, seppur senza successo, le proprie vere credenze per ragioni di desiderabilità sociale non è un caso di bias implicito. E anche se Giorgia avesse anche un bias implicito, il caso esula dagli scopi di questo lavoro. Sia Donald sia Giorgia sono razzisti non conflittuali e non avversi; sono casi in cui gli atteggiamenti espliciti e il bias implicito sono allineati.

Al contrario, se John diventasse consapevole del proprio comportamento discriminatorio dovuto a bias implicito, si sentirebbe profondamente alienato da tale comportamento poiché esso rappresenta per lui una negazione dell'egualitarismo per cui si batte, in cui genuinamente crede.

John rappresenta il banco di prova per testare la nozione di responsabilità morale con riferimento a comportamenti indotti dal bias implicito. Il caso dell'egualitario che discrimina non sembra, dopotutto, una rarità. Come osservano Michael Brownstein e Jennifer Saul, anche se molti fattori concorrono al persistere delle ineguaglianze fra gruppi, dall'eredità storiche, agli stereotipi culturali, a leggi non neutrali, tuttavia la sorpresa è prendere atto del fatto che anche persone che sostengono convinzioni egualitarie e non discriminatorie sono parte del problema, come mostrato da recenti studi in psicologia sperimentale (2016a, 2).

4. Responsabilità per il bias implicito

Poiché gli individui non sembrano avere controllo dei propri comportamenti discriminatori derivanti dal bias implicito, molti studiosi sostengono che non possono essere considerati moralmente responsabili di tali comportamenti (Saul 2013a; 2013b; Kelly e Roedder 2008; Levy 2014). Sulla base di una nozione di responsabilità morale che lega il biasimo all'identificazione dell'agente con i suoi comportamenti (Frankfurt 1971) e alla rispondenza a ragioni (Fisher e Ravizza 1998), tali comportamenti, anche se moralmente deprecabili, non possono essere propriamente attribuiti all'agente che non aveva controllo su di essi.

Vorrei provare a fare un po' di chiarezza e sostenere che, per quanto concordi con questi autori sul fatto che il bias implicito non possa essere oggetto di biasimo, nel senso di *blameworthiness*, e quindi della concezione di responsabilità come *attributability*, credo che una certa interpretazione della concezione di responsabilità come *accountability* possa utilmente essere applicata al bias implicito e dar conto della sua dimensione di ingiustizia, indipendentemente da considerazioni di lode e di biasimo, che non sono indispensabili per la nozione di responsabilità come *accountability*.

Il concetto di responsabilità come *attributability* ha a che fare con il grado in cui un'attitudine o un'azione è espressione dell'agente e dell'identità pratica di chi la compie e misura tale corrispondenza attraverso le condizioni di conoscenza e volontarietà (come discuto nei §§ 4.1 4.2 e 4.3). Al contrario, la responsabilità come *accountability* si concentra sulle conseguenze dell'azione (in termini di causazione e non di espressione agenziale) per le quali si può essere comunque ritenuti responsabili (come discuto nel § 5).

Concludo proponendo una visione moderata sulla responsabilità che si può attribuire per il bias implicito. Sostengo che la fonte della responsabilità morale per il bias implicito moralmente problematico si trova nelle istituzioni politiche e non negli individui. La mia tesi comporta una ridefinizione della condizione di conoscenza (o condizione epistemica) della responsabilità in quanto condizione esterna all'individuo, risultato di condizioni sociali che le istituzioni possono e devono influenzare. Secondo questa prospettiva, ciò che rileva non è che l'agente della discriminazione implicita sia a conoscenza del proprio bias implicito, ma che tale conoscenza sia disponibile nel suo ambiente epistemico. Sostengo che i bias impliciti ingiusti sono quelli non-idiosincratici ma esito di una costruzione collettiva e sociale, e codificati in strutture sociali ingiuste. Tuttavia, sostengo che controllare e modificare il proprio ambiente epistemico non è qualcosa che possiamo normativamente richiedere ai singoli individui, poiché sembra essere eccessiva-

mente esigente per poter esser difeso come un dovere di giustizia (sebbene possa naturalmente essere lodato come un atto di virtù), ma che è ragionevole pretendere dalle istituzioni politiche. Sostengo che, sulla base di questa reinterpretazione della condizione epistemica, è possibile avanzare una interpretazione politica della "responsabilità per" la realizzazione delle adeguate condizioni epistemiche attribuibile alle istituzioni politiche. Segue che ciò che gli individui possono e devono fare e controllare non è tanto il perseguimento di strategie individuali di "de-biasing" (che possono naturalmente essere virtuose, per quanto empiricamente controverse quanto all'efficacia), ma l'azione sulle istituzioni politiche affinché queste ultime modifichino le condizioni sociali strutturali che originano e nutrono i nostri bias.

4.1. Le condizioni della responsabilità morale come attributability

Assunzione centrale nella letteratura sulla responsabilità è che un agente può essere considerato pienamente responsabile nella misura in cui l'agente (i) agisce volontariamente (condizione di volontarietà), (ii) sa cosa sta facendo (condizione epistemica o di conoscenza). Le due condizioni nel loro insieme definiscono il controllo dell'agente sulle sue azioni che è pertanto la chiave della possibilità di attribuire piena responsabilità morale e non solo (*attributability*). Vediamo come il bias implicito influisce sulle due condizioni e rende complicata l'attribuzione di responsabilità.

4.2. La condizione epistemica e il bias implicito

La condizione epistemica (a) prescrive che l'agente non è responsabile per (non) aver fatto x a meno che sia consapevole di (non) aver fatto x.

Nel caso del bias implicito qui discusso, la condizione epistemica non è soddisfatta *ex definitione*: l'agente non sa che si sta comportando in modo discriminatorio e così non soddisfa la condizione (a). Tuttavia, ci sono casi in cui tendiamo ad attribuire responsabilità morale e che pure non soddisfano (a). Tendiamo, per esempio, a ritenere un genitore responsabile per aver dimenticato il proprio figlio neonato in automobile sotto il sole anche se il genitore si era genuinamente dimenticato (e quindi non sapeva) che il figlio fosse lì. Tendiamo a considerare un chirurgo responsabile se dimentica della garza nello stomaco di un paziente anche se nel farlo è genuinamente inconsapevole che la garza sia ancora lì. Attribuiamo responsabilità anche per azioni inconsapevoli perché, come hanno osservato John Randolph Lucas (1993, 52) e Steven Sverdlik (1993, 141), *avrebbero dovuto* sapere. L'ignoranza non è infatti sempre una scusa; a volte è ragione per criticare l'operato dell'agente (Smith 1983, 543).

Quando attribuiamo responsabilità morale (e spesso legale) nei casi di azioni inconsapevoli, la condizione epistemica (a) viene trasformata in una condizione disgiuntiva epistemico-normativa per cui (a) diventa:

(a2) l'agente sa o dovrebbe sapere (o avrebbe dovuto sapere).

Il ruolo del "dovrebbe" in questa formulazione è analogo a quello che George Sher nella sua teoria chiama «operatore deontico» (2009, 72). Inseriamo un operatore deontico nella condizione epistemica affinché quest'ultima possa anche acchiappare casi di azioni inconsapevoli. La condizione disgiuntiva epistemico-normativa recita allora così:

(a2) l'agente non è responsabile per (non) aver fatto x a meno che sia consapevole di (non) aver fatto x o a meno che avrebbe dovuto essere consapevole di (non) aver fatto x.

Ho detto che la condizione epistemica semplice (a) non è soddisfatta dal bias implicito per definizione. Cosa succede con la più ampia versione (a2) nel caso del bias implicito? Per poter rispondere, occorre notare che possiamo interpretare la componente normativa (il "dovrebbe") della condizione disgiuntiva ("sa o dovrebbe sapere") in due modi possibili:

(a2i) una persona ragionevole in quella circostanza saprebbe (o avrebbe saputo);

o

(a2f) una persona che si fosse comportata in modo adeguato in scelte passate saprebbe (o avrebbe saputo).

In (a2i), la conoscenza rilevante è accessibile, ma l'agente la ignora. "Ragionevole" qui è usato nel modo in cui il termine è solitamente impiegato nella letteratura legale sull'ignoranza colpevole (Smith 1983; Hart 1961; Ashworth e Horder 2013, 182). Se l'agente avesse investigato la situazione come avrebbe dovuto, saprebbe (Smith 1983, 544). L'agente è quindi responsabile per il proprio comportamento inconsapevole se l'agente è responsabile della propria ignoranza. Il nostro caso del bias implicito non sembra soddisfare (a2i) perché, come ho detto, sappiamo che essere un agente autoriflessivo, egualitario e ragionevole non è sufficiente garanzia per la prevenzione del bias implicito. Nel caso del bias implicito l'ignoranza non può essere presentata come colpevole perché ciò che si ignora non è un dato accessibile che l'agente ha colpevolmente ignorato. Cosa che, invece, potremmo dire del chirurgo che ha dimenticato la garza nello stomaco del paziente: il chirur-

go "ragionevole" si sarebbe assicurato che il numero di garze estratte dal paziente dopo l'operazione fosse il medesimo di quelle inserite all'inizio come prevede il protocollo, ma il chirurgo non era consapevole che una garza fosse rimasta nello stomaco perché non le ha contate.

Il bias implicito non soddisfa (a2i) perché il bias implicito non è qualcosa che accade solo a chi si comporta in modo irragionevole, ma può capitare ai migliori tra noi.

Nella seconda interpretazione (a2f), la conoscenza rilevante non è accessibile – a differenza che in (a2i) –, e non è accessibile per ragioni che hanno a che fare con il modo in cui l'agente si è comportato in passato. Se l'agente avesse fatto y, z e m in passato, avrebbe minimizzato le probabilità di fare x inconsapevolmente oggi, ma non ha fatto y, z e m. L'evidenza in questo caso non è accessibile perché l'agente non è stato in grado di acquisirla nel tempo. Questo tipo di posizione trova eco nella letteratura sulla formazione del carattere, specialmente di stampo aristotelico.

Su questa linea, George Sher difende quella che presenta come una teoria neo-Humeana della responsabilità che ruota attorno alla formazione del carattere dell'agente. Secondo Sher, chi sbaglia in modo inconsapevole può essere ritenuto responsabile se la sua azione può essere considerata il risultato di un modello (*pattern*) comportamentale cognitivo e affettivo sul quale l'agente avrebbe potuto intervenire prevenendo la cascata causale che ha portato all'atto ingiusto e inconsapevole. L'agente non è intervenuto, per Sher, perché «is prevented by some aspect of his character of belief-system from putting the pieces together» (2009, 21). Per Sher, possiamo attribuire responsabilità morale per azioni inconsapevoli se possiamo ricondurre in un senso rilevante quelle azioni al carattere dell'agente (Sher 2009, 224).

Come ho detto, il bias implicito non soddisfa la prima interpretazione della condizione disgiuntiva (a2i) (agente ragionevole e ignoranza colpevole). Al contrario, sembrerebbe che, in alcuni casi, il bias implicito potrebbe soddisfare questa seconda interpretazione (a2f) sulla formazione del carattere. Si può sostenere, infatti – sebbene non senza controversia nella letteratura scientifica come dirò –, che l'agente sia in grado di mettere in atto una serie di strategie indirette finalizzate a minimizzare l'impatto comportamentale discriminatorio del bias implicito, per esempio esponendosi a modelli controstereotipici di membri di gruppi stigmatizzati. Quindi, si potrebbe forse sostenere che se John, oltre a essere un egualitario sincero, avesse impegnato sufficiente tempo ed energia nell'autocostruzione di un carattere meglio immunizzato dagli stereotipi e standard sociali che nutrono i bias automatici, sarebbe stato in grado di prevenire l'associazione automatica, per esempio, tra finanza e maschio, che lo ha portato a sottostimare il curriculum di Mary per una posizione di broker finanziario. Dunque, in questa interpretazione

di (a2f), il bias implicito sembra candidabile all'ascrizione di responsabilità morale nel senso di *attributability*, nella misura in cui il bias implicito può essere visto come una conseguenza di precedenti azioni o omissioni. Si tratta di una versione della concezione aristotelica di responsabilità (Levy 2005), secondo la quale le disposizioni presenti non sono sotto il controllo dell'agente, ma la formazione delle disposizioni sì.

Tuttavia, diversi studiosi del bias implicito sono scettici sull'efficacia delle strategie indirette per minimizzare il bias implicito (Bargh 1999; Hardin e Banaji 2013). I bias impliciti vengono spesso presentati come molto resistenti agli sforzi di modificarli e metterli sotto controllo (Brownstein e Saul 2016b, 2; Washington e Kelly 2016, 25), ed è stato sostenuto che allertare le persone al riguardo può persino portare a effetti controproducenti, per cui il tentativo di soppressione amplifica l'espressione inconscia del bias (Brownstein 2015; Huebner 2009; Follenfant e Ric 2010; Macrae *at al.* 1994). D'altra parte, visto che il dibattito empirico sul punto è ancora in corso e visto che la questione resta controversa[5], assumiamo, per bontà d'argomentazione, che il bias implicito soddisfi (a2f); vale a dire, che abbiamo un certo grado di controllo sulla formazione del nostro carattere che ci consente di diventare il tipo di persona che è, in una buona misura, immunizzata dalla manifestazione comportamentale del bias implicito.

È sufficiente sostenere che, in presenza di questo potere di formazione del carattere, la condizione (a2f) giustifica l'ascrizione di responsabilità? No, non credo, perché, come ho detto, la condizione (a2f) è una condizione epistemico-normativa (sa o dovrebbe sapere) e la parte che stiamo qui discutendo è appunto quella normativa (cosa avrebbe dovuto fare per sapere). Questo significa che la componente normativa di (a2f) non deve solo essere empiricamente possibile ma normativamente esigibile come un dovere morale il non-soddisfacimento del quale giustifica l'attribuzione di responsabilità morale. La mia impressione è che, sebbene possiamo assumere come empiricamente possibile il controllo sulla formazione del proprio carattere finalizzato a prevenire l'insorgenza del bias implicito, tale forma di autocontrollo e autoformazione sembra essere normativamente troppo esigente. Se da un lato il dovere implica il potere, è anche vero che non tutto ciò che si può fare, si debba anche fare. Penso che la componente normativa di (a2f) nel caso di bias implicito sia eccessivamente esigente e quindi non giustificabile per due ragioni. La prima ha a che fare con l'idea che, secondo un'influente

[5] Ci sono studiosi che difendono una posizione più ottimista secondo cui potremmo, più o meno indirettamente, provare a modificare i nostri bias impliciti (Barden *et al.* 2004; Wittenbrink *et al.* 2001).

tradizione del liberalismo politico, nelle nostre interazioni pubblicamente rilevanti dobbiamo rispettare le persone per quello che sono senza interferire con il loro mondo interiore, incluso il processo di formazione del carattere (Rawls 1993; Carter 2011). La seconda ragione ha a che fare con l'idea che questa interpretazione della componente normativa in (a2f) implicherebbe la verità di una qualche versione del volontarismo doxastico, vale a dire, dell'idea che le persone possono, quando vogliono, modificare le loro credenze. Penso che il volontarismo doxastico sia normativamente inesigibile *qua* impossibile, o normativamente troppo esigente quando possibile, specialmente quando si tratta di stati cognitivi e affettivi non pienamente consci come nel caso del bias implicito. Ciò non esclude che possiamo lodare come virtuosa una persona che riesca a incidere in questo senso sul processo di formazione del proprio carattere, ma non sembra essere materia di giustizia, intesa come ciò che ci dobbiamo gli uni agli altri; materia sulla quale le istituzioni politiche hanno poteri e doveri di controllo e di implementazione. Dunque, se anche avessimo il potere di influenzare il processo di formazione del nostro carattere in modo da diventare agenti senza bias implicito – un "se", come detto, empiricamente controverso – trasformare tale potere in dovere comporterebbe un'intrusività irrispettosa da parte delle istituzioni, e sarebbe eccessivamente esigente nell'assumere una buona dose di volontarismo doxastico.

Posso quindi concludere questo paragrafo dicendo che la condizione epistemica per l'attribuzione di responsabilità morale, nella sua forma semplice e nelle sue due varianti disgiuntive, non può essere soddisfatta dal bias implicito. Passo quindi ora alla seconda condizione standard della responsabilità come *attributability*, quella della volontarietà.

4.3. La condizione di volontarietà e il bias implicito

La condizione di volontarietà (b) può essere interpretata in almeno due modi:

(b1) la capacità di fare altrimenti (Widerker e McKenna 2003);

oppure

(b2) una interpretazione più debole secondo cui la condizione di volontarietà è soddisfatta nella misura in cui l'agente sottoscrive ciò che sta facendo indipendentemente dal fatto che avrebbe potuto fare altrimenti (Radoilska 2015, 2; Fisher e Ravizza 1998).

(b2) include azioni, credenze e atteggiamenti rispetto a cui non avevamo scelta, ma che a seguito di riflessione possiamo avvallare.

(b1) non è chiaramente soddisfatta in caso di bias implicito perché il bias implicito non può essere soppresso o modificato da meri atti volitivi. Questo vale indipendentemente dalla posizione che assumiamo in merito alla controversia sulla possibilità di modificare il bias attraverso strategie comportamentali indirette. Vale indipendentemente perché la condizione di volontarietà (b1) è da intendersi soddisfatta al momento in cui l'azione viene intrapresa: sono responsabile per aver fatto x se potevo fare altrimenti nel momento in cui ho fatto x. Il bias implicito non soddisfa (b1).

(b2) cattura qualcosa di importante in relazione alla dimensione implicita del bias, vale a dire, il fatto che la nostra vita pratica è in misura considerevole modellata da determinate strutture sociali, che di conseguenza formano credenze, atteggiamenti, volizioni e linee di condotta al di là della nostra scelta e consapevolezza (Dasgupta 2013, 235). Nel caso di (b2), sono responsabile per aver fatto x, anche se (b1) non è soddisfatta, ossia non ero in grado di concepire opzioni alternative, e tuttavia approvo x dopo riflessione. (b2) può essere soddisfatta da quei casi di bias implicito che ho chiamato non-avversi, quando cioè gli atteggiamenti espliciti e il bias implicito sono allineati come nel caso di Donald. Donald ha credenze esplicite razziste e bias implicito razzista che, se ne diventasse consapevole, sottoscriverebbe senza problemi. Ma questo non è il caso di bias implicito avverso dell'egualitario John. Se John diventasse consapevole del proprio comportamento discriminatorio da bias implicito, rifiuterebbe il proprio comportamento così che (b2) non sarebbe soddisfatta.

In conclusione, la condizione di volontarietà per l'attribuzione di responsabilità in entrambe le sue versioni non è soddisfatta dal caso di bias implicito qui in esame. Tale considerazione porta quindi a concludere che non possiamo attribuire responsabilità nel senso di *attributability* al bias implicito perché le condizioni standard di tale responsabilità non sono soddisfatte. Ciò potrebbe portarci a concludere con John Bargh su quella che chiama «una conseguenza terribilmente deprimente per la responsabilità morale» (1999, 363) per faccende come il bias implicito. Non è tuttavia detta l'ultima parola e, per dirla, possiamo passare all'altra influente concezione della responsabilità.

5. *Accountability* e bias implicito

L'*accountability* è una concezione di responsabilità compatibile con il funzionamento deterministico del bias implicito. Infatti, se una qualche forma di intenzionalità è necessaria per l'*attributability*, non è così per l'*accountability* che guarda al carattere oggettivo dell'azione commessa. Possiamo quindi procurare un dan-

no ad altri senza intenzione e consapevolezza, e questi altri possono ritenerci responsabili nel senso di *accountability* per non aver di fatto soddisfatto alcuni doveri morali che avremmo dovuto soddisfare (Watson 1996, 262). Per esempio, riteniamo responsabile un genitore che dimentica il proprio bambino neonato in una automobile bollente, anche se il genitore era genuinamente inconsapevole che il bambino fosse lì. Attribuiamo questa responsabilità perché la nostra comunità morale attribuisce uno *status* normativo e certi doveri al ruolo sociale "genitore"; doveri che includono "non dimenticare il figlio neonato in una automobile bollente".

Analogamente, Robin Zheng ha sostenuto che, sebbene non possiamo attribuire responsabilità come *attributability* nel caso di bias implicito in quanto la responsabilità in questo senso è connessa a giudizi come il biasimo indirizzato alla persona come agente morale, non abbiamo questo problema con l'*accountability* perché quest'ultima, e le punizioni che prescrive, sarebbero prive di quella dimensione valutativa-espressiva dell'*attributability* (Zheng 2016). Non mi è chiaro il senso in cui Zheng considera le punizioni come risposte che non esprimono giudizio come se qualsiasi punizione non avesse una dimensione espressiva (Feinberg 1970). Questa è una ragione tra le altre per cui, nella mia prospettiva, la responsabilità come *accountability* per il bias implicito deve essere messa in capo alle istituzioni politiche e non a individui specifici.

Nel prossimo paragrafo delineo il profilo di responsabilità che mi pare plausibile attribuire al caso del bias implicito.

5.1. Responsabilità politica per il bias implicito

Sembra che ciò che ci rimane sia un'interpretazione base di *accountability* priva di tratti di *attributability*[6]. Una versione consequenzialista di *accountability* per cui ciò che conta non è se l'agente meriti di essere biasimato o lodato, ma se l'ascrizione di responsabilità conduce a esiti sociali desiderabili. In questa prospettiva l'ascrizione di responsabilità diventa uno strumento di controllo sociale (Strawson 1974, 2; 20). Una concezione di responsabilità a volte detta *forward-looking*[7].

---

[6] Sembra infatti che anche concezioni miste che combinano tratti di *attributability* e *accountability*, come quella di *answerability* difesa da Angela Smith (2012) (che qui non ho spazio per discutere), finiscono per incontrare problemi simili a quelli individuati nel paragrafo 4.

[7] Thomas Reid e Immanuel Kant sono due esempi dell'idea che possiamo attribuire responsabilità morale se e solo se l'agente se lo merita; e l'agente se lo merita se soddisfa certe condizioni. Thomas Hobbes e John Stuart Mill sono due esempi dell'idea secondo cui, anche se l'agente non meritasse l'ascrizione di responsabilità morale, possiamo giustificare tale ascrizione se le

Penso che questa concezione di responsabilità possa fare al caso nostro se l'onere di questa responsabilità che guarda in avanti viene messo in capo alle istituzioni politiche e non agli individui. Per compiere questo passaggio dagli individui alle istituzioni è sufficiente prendere sul serio ciò che sappiamo sulla fonte sociale del bias implicito moralmente rilevante. Il bias moralmente rilevante si nutre in strutture sociali ingiuste su cui le istituzioni hanno questo tipo di responsabilità. Inoltre, riconoscere la fonte sociale del bias implicito ci consente anche di rifuggire dallo scetticismo morale a cui una visone eziologica meramente psicologica ci condannerebbe. David Wellman, per esempio, osserva che «[i]f bias is ultimately a function of biology and neurology, human actors do not control it. Consequently, they cannot be held accountable for discriminatory behavior [...] [this] might provide the ground for an effective defense against allegations of discrimination» (2007, 50). Tuttavia, il bias implicito moralmente rilevante non è meramente una funzione biologica e neurologica, ma è una funzione della società che segue le tracce di tendenze sociali, stereotipi e ideologie esistenti, di modelli esistenti di privilegio e di svantaggio. Se fosse solo una funzione biologica e neurologica, non saremmo in grado di distinguere i bias moralmente rilevanti, costruiti collettivamente, dai bias individuali e idiosincratici che non derivano dalle strutture sociali in cui uno vive (come il bias implicito contro gli uomini alti in ragione del fatto di aver avuto un padre alto aggressivo e non in ragione di vivere in una società che ospita un diffuso stigma sociale verso gli uomini alti)[8].

Catherine Hundleby ha sostenuto, su questa linea, che il «bias does not constitute a personal moral failing but a *shared* cognitive difficulty» (2016, 257; corsivo originale) perché il bias implicito moralmente rilevante è ereditato dall'ambiente epistemico in cui uno vive. Tamara Gendler e Andy Egan hanno sostenuto che l'intensità del bias implicito è in correlazione con la conoscenza che gli individui hanno degli

sue conseguenze sono giustificabili. Per un esempio contemporaneo della prima posizione, si veda Watson (1987); per un esempio della seconda Smart (1961) o Brandt (1992).

[8] Ci si potrebbe domandare cosa faccia realmente la differenza tra il bias moralmente rilevante e il bias idiosincratico. Se la mia azione affetta da bias idiosincratico contro gli uomini alti concerne le mie funzioni di pubblico ufficiale nella selezione del personale di un ente pubblico, le conseguenze discriminatorie sembrano altrettanto serie di quelle dovute a un bias sessista o razzista. Ciò che fa la differenza sono le diverse fonti dei due comportamenti discriminatori. Nel caso del bias moralmente rilevante, si tratta di un'acquisizione sociale che dipende dall'iniquità del contesto in cui si vive. Nel caso del bias idiosincratico, si tratta di un'acquisizione tutta individuale e psicologica, che crea certamente conseguenze indesiderabili ma paragonabili a calamità naturali, e difficilmente concepibili come faccende di ingiustizia.

stereotipi esistenti, indipendentemente dalla simpatia che provano per tali stereotipi (Gendler 2008a, 2011; Egan 2011). Come osserva Alex Madva, «knowing what the stereotypes are seems to make individuals more likely to act in biased ways» (2016, 191; Correll *et al.* 2002). Su questa linea, Patricia Devine ha sostenuto che il bias implicito riflette la mera conoscenza degli stereotipi culturali esistenti e che ciò è ulteriormente dimostrato dal fatto che egualitari e non egualitari sono vittime di simili associazioni implicite automatiche (1989). Quello che i cittadini possono fare, quindi, è responsabilizzare le istituzioni relativamente ad azioni volte a modificare quegli stereotipi, norme sociali e ideologie che nutrono i nostri bias impliciti. Questa concezione di responsabilità ha lo scopo di isolare e individuare dei doveri di giustizia specifici in virtù dei diversi ruoli istituzionali. Quando diciamo che l'agenzia ministeriale per l'ambiente è responsabile per l'inquinamento dell'aria e dell'acqua, non stiamo ovviamente dicendo che l'agenzia è responsabile di aver inquinato l'aria e l'acqua, ma che è responsabile (ossia ha dei doveri quanto a) del monitoraggio e della prevenzione dell'inquinamento. Naturalmente, il mio approccio porta a doversi chiedere come le istituzioni politiche liberali possano legittimamente intervenire sulle strutture sociali, gli standard e le norme sociali prevalenti; quelle strutture sociali – razziste, sessiste, classiste, eteronormative – con cui entriamo in sintonia inconscia. Questione, questa, che esula dagli scopi e dallo spazio di questo articolo.

6. Conclusione

Ho presentato il bias implicito come un'associazione automatica cognitiva e/o affettiva tra una proprietà stereotipica e un gruppo generalmente oggetto di una qualche forma di stigma sociale. Tale associazione è attivata non intenzionalmente, inconsapevolmente, e talvolta inconsciamente, e quindi influenza in buona misura il giudizio e il comportamento dell'agente al di là del suo controllo.

Ho sostenuto che non possiamo attribuire responsabilità morale nel senso di *attributability* al bias implicito perché ciò che manca nei casi come in quello di John è precisamente quel tipo di rispondenza tra la propria identità e prospettiva valoriale e il proprio comportamento. Ho difeso l'idea di guardare alla concezione di *accountability* e di farlo spostando il focus dagli individui alle istituzioni, e di ritenere queste ultime responsabili per la trasformazione dell'ambiente epistemico e delle condizioni sociali che nutrono il bias implicito.

Il bias implicito viene spesso presentato in letteratura come un problema di malfunzionamento cognitivo; con una lettura meramente psicologica di questo fenomeno sociale, come se fosse, per dirla con Lawrence Blum, una «free floating cognitive distortion» (2016: 161) slegata da ineguaglianze strutturali presenti nella

società[9]. Temo che questo tipo di resoconto del bias implicito rischi di depoliticizzare la questione e di ignorare il tipo di ingiustizia strutturale che crea e nutre il bias implicito. Per di più, è scorretto definire il bias implicito come un difetto cognitivo perché l'associazione automatica del bias implicito può esprimerne generalizzazioni epistemicamente valide e, tuttavia, moralmente problematiche. Per esempio, *ci sono* meno donne matematiche; in alcuni contesti i neri *hanno* più probabilità di essere arrestati. Fare un'associazione automatica tra il background economico-familiare di un bambino e il suo potenziale di successo scolastico può essere un'associazione valida nell'esprimere una generalizzazione e, quindi, una scorciatoia cognitiva di fronte a una gran mole di informazioni, e al tempo stesso, l'associazione è moralmente problematica e irrispettosa (Madva 2016, 194). Come nota Tamara Gendler, «[l]iving in a society structured by race appears to make it impossible to be both rational and equitable» (2011, 57).

Ciò che rende il bias implicito ingiusto è il fatto che sia causato (nei casi moralmente rilevanti) da strutture sociali ingiuste e non, naturalmente, dal fatto che nasca da un presunto difetto del nostro funzionamento cognitivo o psicologico. Sebbene il bias implicito non accada sotto il controllo delle persone, sappiamo che accade, in una certa misura, sotto il controllo delle istituzioni. Il bias implicito, quindi, potrebbe più proficuamente essere visto come un malfunzionamento sociale piuttosto che individuale. Questo implica, come osserva Patrick Shin,

> changing our understanding of discrimination from an agent-centered, moralistic conception to a predominantly psychosocial, diagnostic one […] If unconscious discrimination really is best characterized as akin to passing on an infectious disease, then maybe the law should approach the problem of such discrimination not in the traditional manner of assigning individual responsibility and blame, but much more in the manner of addressing an issue of public health (2010, 101).

Cittadini e istituzioni prendono decisioni quotidiane su chi assumere, chi licenziare, su chi fermare per un controllo di polizia, su quali notizie dare, su quali voti dare, su a chi credere e così via. Tutte queste attività e interazioni possono essere mediate dal bias implicito su cui sembra necessario uno sforzo teorico di comprensione e uno sforzo politico di giustizia. Questo articolo spera di dare un contributo a entrambi.

---

[9] La letteratura si è quindi spesso concentrata su tecniche individuali cognitivo-comportamentali di *de-biasing* invece che investigare un approccio più strutturale (Jacobson 2016; Dixon *et al.* 2012). Oltretutto, l'approccio che guarda solo a ciò che l'individuo può fare sembra lasciarci piuttosto indifesi di fronte alla minaccia deterministica che viene dalle evidenze neuroscientifiche sulla cognizione sociale implicita.

Bibliografia

Arpaly N. (2003), *Unprincipled Virtue: An Inquiry into Moral Agency*, Oxford, Oxford University Press

Ashworth A. e Horder J. (2013), a cura di, *Principles of Criminal Law*, 7th ed., Oxford, Oxford University Press

Banaji M. e Greenwald A. (2013), *Blindspot: Hidden Biases of Good People*, New York, Delacorte Press

Barden J., Maddux W.W., Petty R.E. e Brewer M.B. (2004), "Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes", *Journal of Personality and Social Psychology*, n. 87, pp. 5-22

Bargh J.A. (1999), "The cognitive monster: The case against controllability of automatic stereotype effects", in S. Chaiken e Y. Trope (a cura di), *Dual Process Theories in Social Psychology*, New York, Guilford Press, pp. 361-382

Bargh J. e Chartrand T.L. (1999), "The unbearable automaticity of being", *American Psychologist*, vol. 54, n. 7, pp. 462-479

Bertrand M. e Mullainathan S. (2004), "Are Emily and Greg more employable than Lakisha and Jamal?", *American Economic Review*, n. 94, pp. 991-1013

Blum L. (2016), "The too minimal political, moral, and civic dimension of Claude Steele's 'stereotype threat paradigm'", in M. Brownstein e J. Saul (a cura di), *Implicit Bias and Philosophy. Volume 2: Moral Responsibility, Structural Injustice, and Ethics*, New York, Oxford University Press, pp. 147-172

Bradley C. (1993), "Sex bias in student assessment overlooked?", *Assessment and Evaluation in Higher Education*, vol. 18, n. 1, pp. 3-8

Brandt R. (1992), "A utilitarian theory of excuses", in *Morality, Utility, and Rights*, New York, Cambridge University Press

Brownstein M. (2015), "Implicit bias", *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition) a cura di E.N. Zalta, http://plato.stanford.edu/archives/spr2015/entries/implicit-bias/

Brownstein M. e Saul J. (2016a), "Introduction", in Id. (a cura di), *Implicit Bias and Philosophy. Volume 1: Metaphysics and Epistemology*, New York, Oxford University Press, pp. 1-19

– (2016b), "Introduction", in Id. (a cura di), *Implicit Bias and Philosophy. Volume 2: Moral Responsibility, Structural Injustice, and Ethics*, New York, Oxford University Press, 1-8

Carter I. (2011), "Respect and the Basis of Equality", *Ethics*, vol. 121, n. 3, pp. 538-571

Chaiken S. e Trope Y. (1999), *Dual-Process Theories in Social Psychology*, New York, Guilford Press

Clemons J.T. (2014), "Blind injustice: The supreme court, implicit racial bias, and the racial disparities in the criminal justice system", *American Criminal Law Review*, n. 51, pp. 689-713

Correll J., Park B., Judd C.M. e Wittenbrink B. (2002), "The police officer's dilemma: Using race to disambiguate potentially threatening individuals", *Journal of Personality and Social Psychology*, vol. 83, n. 6, pp. 1314-1329

Correll J., Urland G. e Ito T. (2006), "Event-related potentials and the decision to shoot: The role of threat perception and cognitive control", *Journal of Experimental Social Psychology*, vol. 42, n. 1, pp. 120-128

Dasgupta N. (2013), "Implicit attitudes and beliefs adapt to situations: A decade of research on the malleability of implicit prejudice, stereotypes, and the self-concept", *Advances in Experimental Social Psychology*, n. 47, pp. 233-279

Devine P. (1989), "Stereotypes and prejudice: their automatic and controlled components", *Journal of Personality and Social Psychology*, vol. 56, n. 1, pp. 5-18

Devine P.G., Plant E.A., Amodio D.M., Harmon-Jones E. e Vance S.L. (2002), "The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice", *Journal of Personality and Social Psychology*, vol. 82, n. 5, pp. 835-848

Dixon J., Levine M., Reicher S. e Durrheim K. (2012), "Beyond prejudice: Are negative evaluations the problem and is getting us to like one another more the solution?", *Behavioral and Brain Sciences*, vol. 35, n. 6, pp. 411-466

Dovidio J.F. e Gaertner S.L. (2000), "Aversive racism and selection decisions: 1989 and 1999", *Psychological Science*, n. 11, pp. 319-323

Egan A. (2011), "Comments on Gendler's 'The epistemic costs of implicit bias'", *Philosophical Studies*, n. 156, pp. 65-79

Faucher L. (2016), "Revisionism and moral responsibility for implicit attitudes", in M. Brownstein e J. Saul, a cura di, *Implicit Bias and Philosophy. Volume 2: Moral Responsibility, Structural Injustice, and Ethics*, New York, Oxford University Press, pp. 115-144

Fazio R. (1995), "Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility", in R. Petty e J. Krosnick, a cura di, *Attitude strength: Antecedents and consequences*, Ohio State University series on attitudes and persuasion, vol. 4, Hillsdale, Lawrence Erlbaum Associates, pp. 247-282

Fazio R.H., Jackson J.R., Dunton B.C. e Williams C.J. (1995), "Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline?", *Journal of Personality and Social Psychology*, n. 69, pp. 1013-1027

Fazio R.H. e Olson M.A. (2003), "Implicit measures in social cognition research: Their meaning and use", *Annual Review of Psychology*, n. 54, pp. 297-327

Feinberg J. (1970), "The expressive function of punishment", in Id. *Doing and Deserving: Essays in the Theory of Responsibility*, Princeton, Princeton University Press, pp. 95-118

Fisher J.M. e Ravizza M. (1998), *Responsibility and Control: An Essay on Moral Responsibility*, Cambridge, Cambridge University Press

Follenfant A. e Ric F. (2010), "Behavioral Rebound following stereotype suppression", *European Journal of Social Psychology*, vol. 40, n. 5, pp. 774-782

Frankfurt H. (1971), "Freedom of the will and the concept of a person", *Journal of Philosophy*, n. 68, pp. 5-20

Frankish K. (2016), "Playing double: Implicit bias, dual levels, and self-control", in Brownstein e Saul, a cura di, *Implicit Bias and Philosophy. Volume 1: Metaphysics and Epistemology*, New York, Oxford University Press, pp. 23-46

Gawronski B., Hofmann W. e Wilbur C.J. (2006), "Are 'implicit' attitudes unconscious?", *Consciousness and Cognition*, n. 15, pp. 485-499

Gawronski B. e Payne B.K. (2010), *Handbook of Implicit Social Cognition*, New York, Guilford Press

Gendler T.S. (2008a), "Alief and Belief", *Journal of Philosophy*, vol. 105, n. 10, pp. 634-663

– (2008b), "Alief in action (and reaction)", *Mind and Language*, vol. 23, n. 5, pp. 552-585

– (2011), "On the epistemic costs of implicit bias", *Philosophical Studies*, n. 156, pp. 33-63

Glasgow J. (2016), "Alienation and responsibility", in M. Brownstein e J. Saul, a cura di, *Implicit Bias and Philosophy. Volume 2: Moral Responsibility, Structural Injustice, and Ethics*, New York, Oxford University Press, pp. 37-61

Goldin C. e Rouse C. (1997), "Orchestrating impartiality: The impact of 'blind' auditions on female musicians", *NIBER Working Paper No. 5903*, http://www.nber.org/papers/w5903

Goyal M.K., Kuppermann N., Cleary S.D., Teach S.J. e Chamberlain J.M. (2015), "Racial disparities in pain management of children with appendicitis in emergency departments", *The Journal of the American Medical Association Pediatrics*, vol. 169, n. 11, pp. 996-1002

Greenwald A.G., McGhee D.E. e Schwartz J.L. (1998), "Measuring individual differences in implicit social cognition: The implicit association test", *Journal of Personality and Social Psychology*, n. 74, pp. 1464-1480

Greenwald A.G., Nosek B. e Banaji M. (2003), "Understanding and using the Implicit Association Test: I. An improving scoring algorithm 2, *Journal of Personality and Social Psychology*, vol. 85, n. 2, pp. 197-216

Greenwald A.G., Poehlman T., Uhlmann E. e Banaji M. (2009), "Understanding and using the Implicit Association Test: III Meta-Analysis of Predictive Validity", *Journal of Personality and Social Psychology*, vol. 97, n. 1, pp. 17-41

Hardin C.D. e Banaji M. (2013), "The nature of implicit prejudice: Implications for personal and public policy", in E. Shapir, a cura di, *The Behavioral Foundations of Public Policy*, Princeton, Princeton University Press

Hart H.L.A. (1961), *The Concept of Law*, Oxford, Oxford University Press

Haslanger S. (2012), *Resisting Reality: Social Construction and Social Critique*, New York, Oxford University Press

Hieronymi P. (2008), "Responsibility for believing", *Synthese*, vol. 161, n. 3, pp. 357-373

Holroyd J. (2012), "Responsibility for implicit bias", *Journal of Social Philosophy*, vol. 43, n. 3, pp. 274-306

Holroyd J. e Sweetman J. (2016), "The heterogeneity of implicit bias", in M. Brownstein e J. Saul, a cura di, *Implicit Bias and Philosophy. Volume 1: Metaphysics and Epistemology*, New York, Oxford University Press, pp. 80-103

Huebner B. (2009), "Troubles with stereotypes for Spinozan minds", *Philosophy of the Social Sciences*, vol. 39, n. 1, pp. 63-92

Huebner B. (2016), "Implicit bias, reinforcement learning, and scaffolded moral cognition", in M. Brownstein e J. Saul, a cura di, *Implicit Bias and Philosophy. Volume 1: Metaphysics and Epistemology*, New York, Oxford University Press, pp. 47-79

Hundleby C.E. (2016), "The status quo fallacy: Implicit bias and fallacies of argumentation", in M. Brownstein e J. Saul, a cura di, *Implicit Bias and Philosophy. Volume 1: Metaphysics and Epistemology*, New York, Oxford University Press, pp. 238-264

Jacobson A. (2016), "Reducing racial bias: Attitudinal and institutional change", in M. Brownstein e J. Saul, a cura di, *Implicit Bias and Philosophy. Volume 2: Moral Responsibility, Structural Injustice, and Ethics*, New York, Oxford University Press, pp. 173-187

Jost J.T., Rudman L., Blair I.V., Carney D.R., Dasgupta N., Glaser J. e Hardin C. (2009), "The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore", *Research in Organizational Behavior*, n. 29, pp. 39-69

Kahneman D. (2011), *Thinking, Fast and Slow*, New York, Farrar, Straus and Giroux

Kelly D. e Roedder E. (2008), "Racial cognition and the ethics of implicit bias", *Philosophy Compass*, vol. 3, n. 3, pp. 522-540

Krieger L.H. (1995), "The content of our categories: A cognitive bias approach to discrimination and equal employment opportunity", *Standard Law Review*, vol. 47, n. 6, pp. 1161-1248

Lane K., Kang J. e Banaji M. (2007), "Implicit social cognition and law", *Annual Review of Law and Social Science*, n. 3, pp. 427-451

Lee C.J. (2016), "Revisiting current causes of women's underrepresentation in science", in M. Brownstein e J. Saul, a cura di, *Implicit Bias and Philosophy. Volume 1: Metaphysics and Epistemology*, New York, Oxford University Press, pp. 265-282

Levinson J.D. (2007), "Forgotten racial equality: Implicit bias, decision making, and misremembering", *Duke Law Journal*, n. 57, pp. 345-424

Levinson J.D. e Young D. (2010), "Different shades of bias: Skin tone, implicit racial bias, and judgments of ambiguous evidence", *West Virginia Law Review*, n. 112, pp. 307-350

Levy N. (2005), "The good, the bad, and the blameworthy", *Journal of Ethics and Social Philosophy*, n. 1, pp. 1-16

– (2011), "Expressing who we are: Moral responsibility and awareness of our reasons for action", *Analytic Philosophy*, n. 52, pp. 243-261

– (2014), "Consciousness, implicit attitudes and moral responsibility', *Noûs*, vol. 48, n. 1, pp. 21-40

Lucas J.R. (1993), *Responsibility*, Oxford, Oxford University Press

Machery E. (2016), "De-Freuding Implicit Attitudes", in M. Brownstein e J. Saul, a cura di, *Implicit Bias and Philosophy. Volume 1: Metaphysics and Epistemology*, New York, Oxford University Press, pp. 104-129

MacNell L., Driscoll A. e Hunt A.N. (2015), "What's in a name: Exposing gender bias in student ratings of teaching', *Innovative Higher Education*, vol. 40, n. 4, pp. 291-303

Macrae N., Bodenhausen G.V., Milne A.B. e Jetten J. (1994), "Out of mind but back in sight. Stereotypes on the rebound", *Journal of Personality and Social Psychology*, n. 67, pp. 808-817

Madva A. (2016), "Virtue, social knowledge, and implicit Bias", in M. Brownstein e J. Saul, a cura di, *Implicit Bias and Philosophy. Volume 1: Metaphysics and Epistemology*, New York, Oxford University Press, pp. 191-215

Milkman K.L., Akinola M. e Chugh D. (2015), "What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations", *Journal of Applied Psychology*, vol. 100, n. 6, pp. 1678-1712

Moss-Racusin C.A., Dovidio J.F., Brescoli V.L., Graham M.J. e Handelsman J. (2012), "Science faculty's subtle gender biases favor male students", *Proceedings of the National Academy of Sciences*, vol. 109, n. 41, pp. 16474-16479

Nosek B.A., Greenwald A.G. e Banaji M. (2005), "Understanding and using the Implicit Association Test: II. Method variables and construct validity", *Personality and Social Psychology Bulletin*, vol. 31, n. 2, pp. 166-180

Nosek B.A., Hawkins C.B. e Frazier R.S. (2011), "Implicit social cognition: from measures to mechanisms", *Trends in Cognitive Sciences*, vol. 15, n. 4, pp. 152-159

Oswald F.L., Mitchell G., Blanton H., Jaccard J., Tetlock P.E. (2015), "Using the IAT to predict ethnic and racial discrimination: small effect sizes of unknown societal significance", *Journal of Personality and Social Psychology*, vol. 108, n. 4, pp. 562-571

Payne B.K. (2001), "Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon", *Journal of Personality and Social Psychology*, n. 81, pp. 181-192

Pearsons A., Dovidio J. e Gaertner S. (2009), "The nature of contemporary prejudice: Insights from aversive racism", *Social and Personality Psychology Compass*, n. 3, pp. 1-25

Peters D.P. e Ceci S.I. (1982), "Peer-review practices of psychological journals: The fate of published articles submitted again", *Behavioural and Brain Sciences*, n. 5, pp. 187-195

Phillips A. (1998), *The Politics of Presence: The Political Representation of Gender, Ethnicity and Race*, New York, Oxford University Press

Rachlinski J.J., Johnson S.L., Wistrich A.J. e Guthrie C. (2009), "Does unconscious racial bias affect trial judges?", *Notre Dame Law Review*, vol. 84, n. 3, pp. 1195-1246

Radoilska L. (2015), "Rethinking responsibility: The role of judgement and belief", Paper presented at 'Rethinking Responsibility', Birkbeck College, London, 16 June 2015, http://www.radoilska.com/uploads/2/5/9/0/25906976/radoilska_16_june_.pdf

Rawls J. (1993), *Political Liberalism,* New York, Columbia University Press

Saul, J., 2013a, "Scepticism and implicit bias", Disputatio Lecture 2012, *Disputatio*, V, 37, 243-263

– 2013b, "Implicit bias, stereotype threat, and women in philosophy", in K. Hutchison e F. Jenkins, a cura di, *Women in Philosophy: What Needs to Change*, New York, Oxford University Press, pp. 39-60

Scanlon T.M. (1998), *What We Owe To Each Other*, Cambridge (Mass.), Harvard University Press

Schulman M.D., Berlin J.A., Harless W., Kerner J.F., Sistrunk S., Gersh B.J., Dubé R., Taleghani C.K., Burke J.E., Williams S., Eisenberg J.M., Ayers W. e Escarce J.J. (1999), "The effect of race and sex on physicians' recommendations for cardiac catheterization", *The New England Journal of Medicine*, n. 340, pp. 618-626

Sher G. (2009), *Who Knew? Responsibility Without Awareness*, New York, Oxford University Press

Shin P.S. (2010), "Liability for unconscious discrimination? A thought experiment in the theory of employment discrimination law", *Hastings Law Journal*, n. 62, pp. 67-102

Shiffrin R. e Schneider W. (1977), "Controlled and automatic human information processing: Perceptual learning, automatic attending and a general theory", *Psychological Review*, n. 84, pp. 127-190

Smart J.J.C. (1961), "Free will, praise and blame", *Mind*, n. 70, pp. 291-306

Smith A.M. (2005), "Responsibility for attitudes: Activity and passivity in mental life", *Ethics*, n. 115, pp. 236-271

Smith A.M. (2008), "Control, responsibility, and moral assessment," *Philosophical Studies*, n. 138, pp. 367-392

– (2012), "Attributability, answerability, and accountability: In defense of a unified account," *Ethics*, n. 122, pp. 575-589

Smith H. (1983), "Culpable ignorance", *The Philosophical Review*, vol. XCII, n. 4, pp. 543-571

Stanley D., Phelps E. e Banaji M. (2008), "The neural basis of implicit attitudes", *Current directions is Psychological Science*, n. 17, pp. 164-170

Steinpreis R., Anders K.A. e Ritzke D. (1999), "The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study", *Sex Roles*, vol. 41, nn. 7-8, pp. 509-528

Strawson P.F. (1974), *Freedom and Resentment and Other Essays*, London, Methuen

Sverdlik S. (1993), "Pure negligence," *American Philosophical Quarterly*, vol. 30, n. 2, pp. 137-149

Unkelbach C., Forgas J.P., Denson T. (2008), "The turban effect: The influence of muslim headgear and induced affect on aggressive responses in the shooter bias paradigm", *Journal of Experimental Social Psychology*, vol. 44, n. 5, pp. 1409-1413

Washington N. e Kelly D. (2016), "Who's responsible for this? Moral responsibility, externalism, and knowledge about implicit bias", in M. Brownstein e J. Saul, a cura di, *Implicit Bias and Philosophy. Volume 2: Moral Responsibility, Structural Injustice, and Ethics*, New York, Oxford University Press, pp. 11-36

Watson G. (1987), "Responsibility and the limits of evil," in F.D. Schoeman, a cura di, *Responsibility, Character, and the Emotions*, New York, Cambridge University Press

– (1996), "Two faces of responsibility," *Philosophical Topics*, vol. 24, n. 2, pp. 227-248

Wellman D. (2007), "Unconscious racism, social cognition theory, and the legal intent doctrine: The neuron fires next time", in H. Vera e J.R. Feagin, a cura di, *Handbook of the Sociology of Racial and Ethnic Relations*, New York, Springer, pp. 39-65

White A.A. e Chanoff D. (2011), *Seeing Patients: Unconscious Bias in Health Care*, Cambridge, (Mass), Harvard University Press

Widerker D. e McKenna M. (2003), a cura di, *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, Aldershot, Ashgate

Wittenbrink B., Judd C.M. e Paek B. (2001), "Spontaneous prejudice in context: Variability in automatically activated attitudes", *Journal of Personality and Social Psychology*, n. 81, pp. 815-827

Zheng R. (2016), "Attributability, accountability, and implicit bias", in M. Brownstein e J. Saul, a cura di, *Implicit Bias and Philosophy. Volume 2: Moral Responsibility, Structural Injustice, and Ethics*, New York, Oxford University Press, pp. 62-89