

Implicit Virtue

Susan Stark
Bates College

Keywords: IAT, Implicit Associations, Moral theory; moral psychology; virtue theory.

Many hold that we can be morally assessed only for actions and motivations that are under our control. Recently, however, some researchers have noted that some actions appear to be prompted in part by motivations that are not explicit, but instead are implicit psychological states. These implicit associations are states over which we do not exercise any direct control, though they nonetheless seem to motivate action. But if actions are motivated partly by implicit associations, how can one morally assess them? Individuals do not choose these associations, they are not aware of them, and so it would seem that they are not under their control and thus that they cannot be morally assessed for them. I argue, however, that it is possible to hold ourselves accountable for implicit associations. One moral tradition, virtue theory, holds that individuals can be morally assessed for states that are not fully under their control. Aristotle and some virtue theorists argue that individuals can be morally assessed for their emotions, states that are not under their immediate control. I argue that it makes sense to extend this argument to include implicit associations.

1. Philosophers have long held that we can be morally assessed only for actions that are under our control. Thomas Nagel puts it nicely: “Prior to reflection, it is intuitively plausible that people cannot be morally assessed for what is not their fault, or for what is due to factors beyond their control.” (Nagel, 1979, p. 24) An action is thought to be under our control when the motivation that prompts the action is also under our control. Immanuel Kant underscores the importance of motivation in moral assessment. He holds that it is the motivational states that prompt action, or more precisely, maxims that are the sole object of moral assessment. This Kantian view has an appeal that extends far beyond those who accept Kant’s moral philosophy as true. Kantianism appeals fundamentally to the human aspiration to a just universe. As Bernard Williams eloquently says, the Kantian “conception has an ultimate form of justice at its heart, and that is its allure. Kantianism...offers an inducement, solace to a sense of the world’s unfairness.” (Williams, 1981, p. 21) Kant’s view offers this solace because Kant believed that our motives, and only our motives, were under our control.

Nonetheless, the Kantian aspiration to a just universe appears not to be realized. Evidence for this “bitter truth” (Williams, *ibid*) comes to us, not only from philosophical argument, but also from empirical psychology. Some researchers have found evidence that part of the motivations for certain actions are not explicitly held and reflectively adopted maxims, but instead are implicit psychological states. These non-conscious psychological states, or implicit associations, are states over which we do not exercise any obvious or direct control – indeed, they are states of which we are not even consciously aware. These states

nevertheless contribute to motivating actions, and those actions are morally significant: it has been widely documented, for example, that physicians are more likely to aggressively treat myocardial infarction in white patients than in black patients (Green, 2007). And this difference in treatment results in a difference in outcomes: there is substantial evidence that there are widespread racial and ethnic disparities in health care. As one investigating commission says in its final report: "Racial and ethnic disparities in health care exist even when insurance status, income, age, and severity of conditions are comparable." (Nelson, 2002, p. 666) And these disparities undermine the health and shorten the lives of people of color. Moreover, outside of the context of health care, it has been shown that in circumstances of economic austerity, individuals are more likely to engage in economic discrimination by disproportionately reducing the budgets of organizations that benefit minority and underrepresented groups. (Rudman and Ashmore, 2007) These are just a few examples of the many instances of discrimination that people continue to experience. Put simply: these actions are unjust. They significantly harm people and diminish the justice of society.

These examples of unjust, differential treatment have long puzzled researchers, because the individuals committing the actions appear not to be biased in their explicitly held beliefs. It is not until researchers look to the level of non-conscious mental states, and specifically to implicit associations, that they can offer a plausible explanation of how these prejudicial actions can be committed by people who appear not to be biased in their avowed beliefs.

But now thinking back to the view articulated by Nagel and commonly held, if these actions are motivated by implicit associations, how, can one morally assess them? Individuals do not choose their implicit associations, they are not aware of them, and so it would seem that

they are not under our control. On the basis of this lack of awareness and control, many would conclude that we cannot be morally assessed for these associations.

I argue, however, that despite appearances to the contrary, it is possible to hold ourselves accountable for our implicit associations. There is one moral tradition, virtue theory, which holds individuals morally accountable for states that are not fully under their control. Virtue theorists argue that individuals can be held accountable for their emotions, which in the moment that they experience them, are not under their control. As Aristotle says, "...the man who feels fear or anger is not praised, nor is the man who simply feels anger blamed, but the man who feels it in a certain way." (Aristotle, trans. W.D. Ross, 1925, 1105^b33) To many this quote seems puzzling, or even mistaken, because Aristotle seems to hold both that we are and at the same time are not praised for the emotions we feel. This is because it is true both that we can, and also cannot, control our emotions: we cannot control whether we feel fear or anger, but can control the manner in which, or the extent to which, we feel them. Furthermore, I shall argue that this passage contains an important insight, and one that can help us make sense of the equally puzzling idea that we can hold ourselves accountable for states that we cannot, in the most plausible sense of the term, control.

2. The current understanding of one set of scholars is that individuals, in addition to possessing the standard repertoire of psychological states (including beliefs and desires), are also in possession of implicit associations. In a manner reminiscent of Hume's association of ideas, someone has an implicit association when she links two distinct traits, because she perceives that the traits have often occurred together. Those who study implicit bias thus

measure the ease with which subjects link the evaluative and descriptive concepts: if a subject links them more quickly, the researcher explains this by appeal to the subject's implicit association between the two concepts.

To study implicit associations, researchers have designed a number of tests, known collectively as Implicit Association Tests, (<https://implicit.harvard.edu/>) which measure how quickly a subject sorts words into categories. In one such test, researchers study a participant's normative associations (good and bad) with the race-concepts, 'black' and 'white.' The Implicit Association Test (IAT) asks participants to sort words that flash on the computer screen into categories. So in the first of several sequential tasks, the participant will sort words that appear in the center of the computer screen (joy, love, peace, agony, terrible, horrible, etc.) into one of two categories: "Good" (on the left hand side of the screen) and "Bad" (on the right). Then, in the next task, the participant will sort pictures of faces, some African-American, and others European-American. The participant again sorts them into two categories, this time labeled, "African-American" or "European-American." The next task randomly pairs the evaluative concepts with the descriptive ones. So the choice on the left might read: "African-American or Good" and the one on the right would read: "European-American or Bad." Now both words and images (from the earlier sorting tasks) would appear serially in the center of the screen and the participant would sort them into their appropriate categories. And then in the final task, the pairings would switch: so the choices would then read: "African-American or Bad" and "European-American or Good". And once again the participant would sort the words and images in the center of the screen into their appropriate categories. With each task the participant is instructed to go as fast as possible while making as few mistakes as possible.

The IAT has had hundreds of thousands of participants. These participants have been from many demographic groups in more than 35 countries. Together they have completed more than 4.5 million sorting tasks. (Ibid) In the IAT concerning attitudes about race, researchers have found that most test-takers display some implicit association between European-American and good and between African-American and bad. For some people this association is stronger than others. Specifically, researchers have found that 70% of subjects who completed the race-based IAT exhibit some automatic association between European-American and good. (Only 12% of subjects exhibit an automatic preference for African-American and good and 17% display little to no automatic preference for either race.) Between 75 and 80% of self-identified whites and Asians exhibit some automatic preference for European-American and good. (Ibid) Furthermore, people from all social backgrounds have implicit associations or biases: “from young to old, male to female, Black to White, and conservative to liberal, implicit biases are not held by a select few but are readily observed among all social groups.” (Nosek, Banaji and Greenwald, 2002, p. 112)

The IAT measures the amount of time it takes for subjects to sort the words and images into their appropriate categories. The associations measured by the IAT are considered to be biases because researchers hypothesize that a subject who strongly associates European-American with good (that is, who has that bias) will more quickly sort words and pictures into the category “European-American or Good.” Furthermore, in a subject who has the implicit association between European-American and good it will be more difficult, and thus take more time, to sort words and into the categories, “European-American or Bad,” and “African-American or Good.”

In addition to these implicit preferences, most social groups also show *explicit* preferences for their own groups. For instance, both white Americans and young people show strong *explicit* preferences for their own groups, or ingroup preferences (these are measured by their responses to a questionnaire). Black Americans and elderly people similarly show strong *explicit* preferences for their own groups as well. However, most black Americans, like white Americans, exhibit an implicit preference for white Americans. In other words, they display an implicit association between European-American and good, just as most white Americans do. It is notable that there is a divergence between the explicit beliefs of black Americans and their implicit associations: they show a strong explicit preference for black Americans, but nevertheless show an implicit preference for white Americans. Researchers speculate that implicit associations reveal the “deep influence of the immediate environment and the broader culture on internalized preferences.” (Nosek, et al, *ibid*) Because the United States culture exhibits a pervasive preference for white over black, the members of the culture, whatever their racial background, exhibit a similar preference. In the same way, because the United States culture exhibits a preference for young over old, both young people and the elderly share an implicit association between young and good. In the age-based IAT, researchers found that over 80% of subjects who took the test prefer young people to old. So as with African-Americans and other black Americans, the elderly experience a dissociation between their explicit preference for their own group and an implicit preference for young people.

There are many revealing examples of implicit biases in the IAT. For instance, in another sorting task, participants are asked to sort names (“Ben,” “John,” “Daniel,” “Julia,” “Michelle,” “Anna,” etc.) associated with gender into two categories, “Male” and “Female.” They then sort

words associated with family (“home,” “parents,” “children”) and career (“management,” “professional,” “corporation”) into two categories, “Family” and “Career.” Researchers then combined the categories: so participants sort the words and names into one of two categories: “Male or Family” and “Female or Career.” And then as with the other IATs, the pairings would switch to “Female or Family,” and “Male or Career.” Researchers found that 24% of subjects exhibit a strong association between female and family and between male and career; 32% of subjects exhibit a moderate association between female and family and male and career; and 20% of subjects exhibit a slight association between female and family and male and career (a total of 76% of subjects exhibit some association between female and family and male and career). 17% of subjects exhibited little to no association at all, while only a total of 6.3% of subjects combined exhibited an association, of any strength (strong, moderate or slight) between male and family and female and career. (<https://implicit.harvard.edu/>)

3. These findings are important on their own because they may help us to better understand certain forms of injustice, such as racism, sexism and ageism. Justice is a central human good and thus understanding and overcoming all obstacles to a just society (including prejudicial implicit attitudes) is, in the same way, a requisite human good. But the IAT is also important because there also appears to be a correlation between implicit attitudes and certain sorts of actions. Despite the fact that individuals are not aware of their implicit associations and do not choose to have them, these associations nonetheless seem involved in motivating actions. At the very least, the research seems to show that the IAT is a better predictor of certain sorts of actions than are a person’s explicit associations or beliefs. One study, for

instance, examined the existing disparity in the treatment of myocardial infarction (a kind of heart attack) for black and white Americans. (Green, et al) White Americans are up to twice as likely to receive thrombolytic therapy for heart attack as are blacks. In this IAT study, 393 internal medicine and emergency medicine residents completed three tasks: first they read a vignette describing a 50-year-old male patient who presented to the emergency department with chest pain. Part of the vignette included an electrocardiogram indicating myocardial infarction. It also included a photograph: some of the physicians received a photograph of a black face and others received a photograph of a white face. After reading the vignette, participants rated the likelihood that the chest pain was caused by coronary artery disease, and said whether they would treat the patient with thrombolysis. In the second task, participants answered several questions about whether they preferred black or white Americans, and whether they found black or white Americans to be more cooperative with medical treatment. These questionnaires assessed explicit bias. And in the final task, participants completed three IATs designed to assess implicit racial bias.

The researchers found that in their explicit beliefs, the participants expressed an equal preference for black and white Americans; they felt equally warmly toward black and white Americans; and they found black and white patients to be equally cooperative with medical treatment. Regarding their implicit preferences, however, participants showed a moderate to strong association between white and good, and between black and bad. Finally, regarding their actions toward the patient represented in the vignette, physicians prescribed thrombolysis substantially more often for the white patients than for the black patients: “58.2%

of physicians were very likely to offer white patients thrombolysis versus 42.7% for black patients.” (Green, et al, p. 1235)

What this study shows is that explicit beliefs do not predict whether the physicians will be more likely to prescribe thrombolysis; but implicit associations do: “as the degree of antiblack bias on the race preference IAT increased, recommendations for thrombolysis for black patients decreased.” (Ibid) These prejudicial actions are better predicted by the IAT than by the explicit beliefs. Now certainly it is true that some of the individuals with a strong association between black and bad nonetheless prescribed thrombolysis for their black patients; similarly some of the individuals with a strong association between white and good failed to prescribe thrombolysis for their white patients. Correlation does not imply causation. Nevertheless, the performance on the IAT was a better predictor of the prejudicial actions than was the explicit belief assessment.

A second study (Rudman and Ashmore), this time outside of the context of medicine, examined whether either performance on several race-bias IATs or answers to explicit questionnaires about race were correlated with discriminatory actions (one arm of the study examined an individual’s reports about past discriminatory actions, such as telling a racist joke, or using verbal slurs, or even physically harming someone because of their race; the other arm of the study examined making budget cuts to Jewish, Asian and Black student organizations at a University as opposed to cutting the budgets of organizations like the drama club and the school band). Researchers again found that little or no prejudice was evident in the explicit measures. Performance on the IAT, by contrast, showed a strong correlation between white and good/black and bad; and moreover, the stronger the association between white and

good/black and bad the more likely was the individual to report a history of past discriminatory actions, and the more likely was the individual to engage in economic discrimination (cut the budget of the Jewish, Asian or Black student organizations rather than the race-neutral organizations). (Rudman and Ashmore) So again we see that it is the performance on the IAT rather than the response to a questionnaire that better predicts the discriminatory actions some individuals may take.

This finding may help to explain the persistence of prejudice and discrimination all the while most individuals believe that they are not racist, sexist, or homophobic. But this finding may also be important for the moral theorist: some moral theories, including Kantianism and certain versions of utilitarianism, hold that the motivations for morally assessable actions are explicitly held beliefs. It looks, however, as though at least some morally assessable actions are motivated not by explicit beliefs but instead (in part) by implicit associations. So if individuals are to be held morally accountable for their prejudicial and discriminatory actions, a moral theory must be able to assess the motivations that give rise to those actions: namely, implicit associations.

I suspect that virtue theory is better able to assess these implicit associations than are utilitarianism and Kantianism. Here, however, I focus on explaining how it is that virtue theory can regard individuals as responsible and assessable for implicit associations. Virtue theory can offer an explanation for how implicit associations might motivate actions that are nonetheless morally assessable, because implicit associations are, in several important and notable ways, similar to the ordinary virtue theoretic motivations that motivate actions that the virtue theorist is able to morally assess.

4. Implicit associations are connections or links between descriptive concepts and evaluative ones. Moreover these associations operate “outside of [the actor’s] conscious awareness.” (Banaji, Lemm, and Carpenter, 2001, p. 134) The individual is prolific at forming these associations, whether he wants the associations or is aware of the associations or not: his “...mind is constantly gathering information from the social environment and creating, revising, and reinforcing associations in memory.” (Smith and Nosek, 2010, p. 803) Advertisers, for instance, have long exploited these associations. They know that individuals can be made to have positive associations with Coke rather than Pepsi (or vice versa), even if the individuals have never tasted either Coke or Pepsi. But it is not only with expensive advertisements that individuals form these unconscious associations. Individuals also form positive and negative associations constantly from every day experiences: as we have seen, many individuals in North America in 2012 seem to associate young people with good and old people with bad, women with nurturing and home, men with career and professional life, white race with good and black race with bad, and so on.

How can these prolific implicit associations be subject to moral assessment when they are not chosen and are outside of the actor’s conscious control and awareness? Indeed, as we saw earlier, for hundreds of years, it was commonly held (by philosophers, psychologists, and lay people) that for an action to be the subject of moral assessment, it must be in our control to perform it or not perform it. And in order for the action to be under our control, the motivation must, in turn, be under our control. Again, Bernard Williams explains this view nicely: “Anything which is the product of happy or unhappy contingency is no proper object of

moral assessment and no proper determinant of it, either.” (Williams, p. 20) So how can people be morally assessed for actions that are not their fault, that are due to influences that are not under their control and that are not even part of their conscious awareness?

Developing a credible theory of moral responsibility in the face of moral luck is one aspect of this problem. Thomas Nagel (1979) takes up one aspect of this question, famously concluding that the so-called problem of moral luck is a problem without a solution. Nagel may be right about the global problem of moral luck and moral responsibility. Nevertheless, I want to urge, it may still be possible to get some traction on the narrower issue concerning the matter of moral responsibility and the IAT.

First, we should note that, the process of coming to have implicit associations is the process of having a particular perceptual awareness of reality: advertisers show us images of individuals enjoying Coke and doing things we take to be pleasant (surfing, skiing, going on a date with an attractive person, etc.), and we come to associate Coke with good, we perceive coke as good. In the same way, individuals repeatedly see women in nurturing roles and men in the work force (on TV and in “real life”) and come to associate women with nurturing (or perceive women as nurturers) and men with work-outside-the-home (perceive men as professionals).

Now certainly, it need not be true that women are more nurturing and men are more career-oriented, or that young is good and old is bad, or again that white is good and black is bad, for these associations to take hold in individuals. It simply needs to be the dominant story told by the individual’s culture. Consider for instance the cultural appraisal of youth as more valuable than the aged. In North American culture, many people associate aging with death,

loss of function, loss of autonomy, isolation, loneliness. In the same way, North Americans associate youth with energy, freedom, life, passion, community and connection. So it is not surprising that many individuals in North America show an implicit preference for young over old. It is possible, by contrast, for a culture to associate age with wisdom, seniority, respect and knowledge. And youth, in this culture, might be associated with naiveté, uncertainty, and inexperience. Such a culture might find its members valuing age over youth (having, that is, an implicit association between age and good, youth and bad). So it is the cultural view of youth and age (the norms and ideals associated with these two descriptors) that in part determines the implicit associations of its members. It need not be some fact about youth and age: thus it need not be true that being young is more valuable than being elderly; rather it is simply what a culture takes to be true. In the same way, it is not true that Coke tastes better than Pepsi – indeed, some people prefer the taste of Coke and others the taste of Pepsi (I like Pepsi better myself). These are simply stories told by our culture in many forms – in advertising, in media, in boardrooms, courtrooms, on the assembly line, in job interviews, etc. It may be that as these stories are told, individuals perceive them and absorb them into their implicit associations. And over time, individuals act on these stories and perceive them in many different contexts. Because of this, eventually, these associations may become habitual ways of seeing and acting in the world.

Next, I want to urge that moral perception functions in much the same way. From when we are tiny babies we perceive that certain experiences and actions cause pain and others cause pleasure. These early experiences of aspects of the world as pleasurable or as painful begin to construct our understanding of what has moral importance and what does not. We

come to see, through these experiences that some things are to-be-promoted and others are to-be-avoided. The moral story (the rules and injunctions, requirements and prohibitions, values, norms and ideals) that we are taught comes not only from advertising and media, but in a more primal way, comes from our parents, our caregivers, our teachers and our peers. It comes also from the norms and values of our culture and from our own capacity for reflection. All of these things together—our culture, our caregivers and teachers, our reflections—teach us what typically has moral significance (skinned knees, unfairness, and helpfulness) and what typically does not (finger paint, sneakers and orange juice).

As Aristotle argued, learning to see what has moral significance is a process: it is a process of internalizing the good. The way we do this, according to the virtue theorist, is by practicing virtuous actions. Initially, most of us do morally good deeds because we are told to: we share our toys, we avoid hurting others, not because we have internalized the good, but because we empathize with someone who is hurt, or because we are emulating our caregivers, or because we want to please them, or want to avoid punishment for transgressing. Doing these actions is practice for becoming virtuous. Practicing these good actions cultivates a way of perceiving the world: we come to see certain actions as harmful (throwing sand in someone's face), we come to see certain words as hurtful (teasing) and we come to see certain actions as kind or helpful (playing with a lonely child on the playground). We see these actions in this way, in part, because we have begun to cultivate a moral sensibility: we have begun to affirm the moral values contained in this way of seeing the world. Our perception builds on natural sympathies and also transforms some of the learner's feelings: through emulating virtuous people, through seeing the world as they do, the novice comes to feel that her actions are

good. She enjoys the actions and loves doing them. And this development and transformation of her emotions becomes habitual as this way of seeing the world takes root in her character. Moreover, these ways of seeing and feeling are, according to the virtue theorist, the everyday moral motives that prompt the ordinary person to act in morally good ways.

Furthermore, the virtue theorist argues that these feelings and perception, these motives are not chosen in the moment at which they motivate our actions. Miles Burnyeat argues that our moral feelings cannot be chosen in a moment, just as our feelings about non-moral goods cannot be chosen in a moment. Burnyeat argues that you cannot choose in a moment to love to ski or to find espresso delicious or see that a Picasso painting is beautiful. These feelings, these acquired tastes, must be cultivated over time. Burnyeat further argues that each of these senses (moral or non-moral) must be cultivated over time, through habituation and experience. Coming to love espresso involves drinking it over time, initially with lots of sugar and milk, until the novice espresso drinker cultivates a taste for it. The same goes for skiing, viewing art, and moral virtue. The moral novice emulates the more virtuous people around him because by practicing helpful actions, courageous actions, temperate actions, and so on, he will come to internalize a love of these actions. In the beginning the moral novice may not enjoy helpfulness – he is self-centered or lazy or begrudging, unwilling or reluctant to help. But he helps out of a desire to please a role model or to avoid punishment or because he finds in himself a glimmer of empathy for someone in need of help. As he helps he feels a bit of joy in doing the helpful action. The value of helping for its own sake thus begins to grow in his emotions. And it is the performance of the actions, the practice of doing them, that enables the learner to come to love the actions and love the good in them. (Burnyeat, 1980)

Moral virtue requires the virtuous person to do the right actions and feel the right emotions, as these things are called for by particular moral situations. But how can the virtuous person control, in the heat of a moral situation, which emotions she feels? Aristotle acknowledges that we cannot choose, at a given moment, to feel fear or anger or joy; nevertheless we can choose over time to cultivate those feelings so that the next time, or a few times down the road, when certain feelings are called for, or when we are called upon to love the good (whatever it is) those feelings will be ready and waiting to serve us. Our emotions, then, are chosen in the sense that they are cultivated over time. Our choices are determined by how we are taught and how we take up that teaching. (Sherman, 1995) And because we can cultivate these states through our moral development, we can exercise some control over them. The virtue theorist thus regards the individual's way of seeing the world and the feelings associated with it as subject to moral assessment.

We choose these feelings over time in part because, as Burnyeat notes in developing Aristotle's view, practicing virtuous actions engenders both an affective and a cognitive shift: the moral learner comes to love the good and in so loving it, she comes also to know in a robust and full sense, that these actions are good. Thus the pinnacle of moral development comes when the moral learner has transformed not only her emotions through doing virtuous actions, but also has transformed her cognitive moral understanding. The virtuous person recognizes the fullest extent of the goodness of virtue: she has internalized it or taken it to heart that these things are good.

At this point it may be objected that many of us are not aware of our moral feelings. This is certainly true. Despite the fact that the virtuous person, the moral expert, has a

complete knowledge of the goodness of virtue, it nonetheless remains true that the young moral learner may be unaware of her moral feelings at any given moment in time. For many years during our moral development, it is largely the influence of our teachers and caregivers that enables these feelings and this way of seeing to grow in us. It may not be until we are teenagers or older that we can come to be aware of our moral feelings. For instance, a young child cannot say, “I feel indignant about that injustice;” though a teenager can. The young child does not yet know the difference between anger and righteous indignation and she also lacks a developed sense of injustice, though she certainly has a clear sense of when something is, to her, “Not Fair!” The teenager or young adult, on the other hand, is beginning to understand these more advanced concepts (indignation) and make these more sophisticated distinctions (the difference between “it didn’t go my way” and “it is unfair” for instance). The young adult, in the same way, can endorse the project of cultivating a certain moral sense in our character. It can thus take many years for us to be aware of the feelings we are cultivating in becoming morally virtuous. And some of us, due to our early influences, may be aware of our moral feelings in only a rudimentary way at any point in our lives. Others of us, perhaps many of us, however, do aim to cultivate our moral emotions through our moral development.

We see, then, that some virtue theorists do hold that our moral motivations may lie, at least largely, outside of our articulate, conscious awareness, and may not be subject to immediate choice and control. For some Aristotelian virtue theorists, a motivation need not rise to the level of an explicit belief to motivate morally assessable actions. Nevertheless, the virtue theorist argues that over time we can choose, become aware of, and control our emotions and our love of the good. Together our emotions and love of the good constitute many of our

ordinary virtue-theoretic moral motivations. We control our emotions and love of the good, over time, by listening to the teachings of the moral exemplars in our lives; by choosing to notice and attend to morally salient particulars; and by practicing virtuous actions. In our moral development, in honing our keen moral perception, and in doing the actions that are morally required, we cultivate our emotions, so that over time, they line up with the good and motivate morally correct, and hopefully one day, virtuous, actions.

Some virtue theorists have embraced this account, especially because it tells a powerful story about how individuals can overcome the bias that has found its way into their moral persona. Rosalind Hursthouse (1999), for instance, emphasizes the important role of moral education and has argued that racism is a paradigm case of bad moral training. Though she does not discuss the IAT per se, she nonetheless argues that the virtue theorist offers a compelling view of how we can unlearn the effects of bad moral training (these effects include racism, sexism, ageism, homophobia, etc). Furthermore, Hursthouse argues that individuals have a powerful motivation for unlearning these biases: it is the love of justice. She asks, “Why do those of us who had racism inculcated in us think that we must strive, and continue to strive, to undo the effects of that upbringing? Not because we think it will make it easier for us to do what is charitable and just, but because we think it will make us better people, more charitable and just than we are at present.” (Hursthouse, p. 117) As Hursthouse argues, a love of the good can be a powerful force to impel individuals to overcome their unjust implicit associations and work toward becoming more just individuals and creating more just societies.

5. We have seen that the virtue theorist provides moral agents a framework within which agents can morally assess motivational states that are not under their immediate control, not within the agent's awareness at the time when they first begin to motivate her, and finally not chosen in the moment they are experienced by the agent herself. Now further notice that the affective shift that occurs for the moral novice with the ordinary virtue-theoretic motivations is the same transformation that occurs with implicit associations. As we have seen, individuals do not choose to have positive feelings about helping behavior. Their parents, teachers, and role models require helpful behavior and an appreciation of helping begins to grow in them. The more individuals help in the future, the more the good feeling grows. In the same way, individuals do not choose their implicit associations: they do not choose to associate European Americans with good, African Americans with bad; they do not choose to associate women with home and family, men with career and business; they do not choose to associate young with good, aged with bad. These associations come to them in part because of the values and norms of their cultures. These norms imprint upon individuals and seep into the fabric of their lives.

Now it will be noted that part of the path of moral development for the virtue theorist involves the recognition of the moral feelings and the recognition of the moral values associated with them. We saw that practicing virtuous actions is meant to engender a cognitive shift as well as an affective one. Cultivating these positive feelings enables the moral novice to learn something cognitive about the moral virtue (or again, in the case of non-moral skills, about the value of skiing, espresso, and Picasso paintings: he learns that skiing is enjoyable, that

espresso tastes good, and that the Picasso is beautiful). Regarding moral virtue, the novice learns that helping others is valuable for its own sake.

However, it might be noticed, these examples of skiing, espresso-drinking, Picasso paintings, meant to be explanatory of how we become virtuous, are examples in which our explicit beliefs are transformed: individuals come to know explicitly that skiing is enjoyable, that espresso tastes good, and that the Picasso is beautiful. And in the moral case, individuals come to know that helping is good, loveable, and virtuous. But implicit associations are, as we have seen, implicit: they are associations that transform the individual's feelings, but do not change her cognitions, her explicit beliefs. This seems to be an important disanalogy.

In response, however, we can see that even though implicit associations may not transform beliefs, the beginnings of the two processes look identical. To begin with the non-moral case, in the early days of learning to ski, most individuals do not believe that skiing is an enjoyable activity. Rather, they find it cold and unpleasant. It is hard work and sometimes terrifying to strap long wooden boards onto uncomfortable, stiff, boots and point them down (what looks to be) a sheer, icy slope. So in the early days, most find skiing to be almost entirely unpleasant. But then over time, with some glimmers of pleasure here and there, some successes and lots and lots of practice, some both come to love the activity and also to know that it is enjoyable.

The same is true of moral virtue. In the beginning, as we have seen, individuals often do not find helping, or sharing, or resisting our baser instincts to be pleasant. It is hard and unpleasant, and often, individuals do not want to do it. But with practice, and again with glimmers of pleasure here and there (even very young children can and do act empathetically),

individuals come to enjoy helping and sharing, and similarly become slower to anger or retaliate, so that they begin to love the good and begin to become virtuous. These activities come to be seen as pleasant, loveable, good and virtuous. Practicing virtue, as with skiing, espresso-drinking and viewing Picasso paintings, transforms the individual's beliefs, so that she comes to know something cognitive (some know-how) about helping. She comes to know how to help and loves helping.

The process of coming to have implicit associations is similar to the early part of the process of coming to be virtuous. Implicit associations are like the associations of feeling individuals have at the beginning of learning to ski or learning to be virtuous: they are non-conscious and have not (yet) reached the level of explicit belief. Now it might be, unfortunately, that with enough time and enough practice committing prejudiced actions, individuals would experience a transformation of their beliefs, so that they would offer racist or sexist responses to questions about their beliefs or temperature readings of feelings of warmth or coldness toward the groups in question. Indeed, we know that some individuals do experience a transformation of their explicit beliefs about race or gender: they are openly racist and sexist.

However, it might be that some individuals have a dissociation between their implicit associations and their explicit beliefs precisely because the associations in question are biases. Many individuals are committed to seeing themselves as unbiased, and so are ashamed of any biases they may suspect. For these individuals, the dissociation between their implicit associations and their explicitly neutral beliefs might best be explained by interference from a desire to live an unbiased life – this desire masks or hides the implicit association from the

individual's conscious awareness. Many individuals thus have a strong counter motivation for keeping these attitudes at the implicit level, thus keeping them hidden from their conscious awareness.

But even if the process of coming to have implicit associations is identical to the early stages of the process of coming to have ordinary virtue-theoretic motivations, a more worrisome concern is that implicit associations cannot be changed once they have infected an individual's life. So even if the virtue theorist can explain how it is that individuals come to have the associations, if they cannot be changed, then it would still seem that individuals cannot be morally assessed for them. Indeed, Aristotle himself might have been skeptical that the associations could be changed: "It makes no small difference, then, whether we form habits of one kind or of another from our very youth; it makes a very great difference, or rather all the difference." (Aristotle, 1103^b22-25) On this view, implicit associations are unfortunate but unchangeable bad habits of feeling. They make it impossible for individuals to be virtuous and impossible to lead the best human life. They are the result of living in an unjust society. And again, as Aristotle reminds us, not only does it matter what sort of upbringing individuals have, but it also matters what sort of society they live in: an individual cannot live the best human life in an unjust society. The existence of oppression in society makes everyone worse off.

Though I agree that oppression and injustice tarnish the lives of all in an unjust society (though not all equally), there is research that shows that it is possible to change implicit associations. One study, for instance, shows that the association between African American and bad can be significantly reduced after exposure to images of well-liked African Americans. (Dasgupta and Greenwald, 2001) And this finding, too, can be interpreted in an Aristotelian

way: this positive experience is a sort of habituation. It begins or continues the process of transforming an individual's feelings to associate African-American with good, a process that tracks the important human good of justice.

We are now in a position to see that implicit associations share several important things in common with ordinary virtue-theoretic moral motivations: both sorts of motivations involve a non-cognitive transformation. Ordinary virtue-theoretic moral motivations, in their more advanced state (that is, for the more morally developed individual) also begin to involve a cognitive shift that is typically absent for implicit associations. However, the motivation for morally assessable actions, done by the beginning student, involves affective states that are analogous to implicit associations. Implicit associations and virtue-theoretic motivations also share in common the possibility of changing these motivations over time. Neither sort of motivation can be changed or chosen in the moment: individuals cannot choose to love the good at one particular moment in time; they cannot choose to associate white with good nor can they choose to stop this association at one moment in time. However, we saw earlier that individuals can cultivate their moral feelings over time so that they come, eventually, to have them. The same goes for implicit associations: through positive experiences implicit biases can be mitigated and perhaps even overcome, over time; through negative experiences or biased role models they come to be more engrained. And an individual's choices can to some extent dictate which sort of experiences we have: individuals can choose to associate with people who further ingrain their implicit biases, or they can choose to associate with those who will help them overcome them. We can thus see that virtue theory, because it already provides for the

development and moral assessability of feelings that are unchosen in the moment, can similarly explain and morally assess implicit associations and the actions they motivate.

But even still, one might worry about this view that there remains a crucial disanalogy between ordinary virtue-theoretic moral motivations and implicit associations: of the former the individual is aware as they are motivating her, of the latter, the individual is not aware as they are motivating her. So when the moral novice is learning the virtue of generosity, though she may find it difficult to give away some money or a treasured item, she may nonetheless find a small glimmer of pleasure in doing so. This feeling of enjoying the generous action, crucially, of which the actor is aware, is then able to grow, through future coming-to-be-generous actions, until she more fully establishes the virtue generosity in her character. On the other hand, when the emergency department physician prescribes thrombolysis for her white patients more often than for her black patients, she may have no idea, first, that she is prescribing thrombolysis more frequently for one population than for the other, and second, she may have no idea that this difference is due to her own implicit attitude about the patient's race. Indeed, if questioned about it, she would likely offer medical justifications for the difference in her treatment patterns. This is an important point: for how can we possibly work to change an association we have, if we do not even know we have it. It seems again that we are unable to hold people accountable for the implicit associations, this time because they are attitudes of which individuals are unaware.

Notice, however, that not all physicians are unaware of these racial differences in prescribing patterns. The physicians who participated in the thrombolysis study are no longer unaware of this difference; nor are the physicians who read the published report of this study.

Indeed, hundreds of thousands of people have now themselves taken many IATs (millions of them all together), and they, very likely, are aware of these differences, or could easily be made aware of them (for instance, by reading the abstract for the published report of the thrombolysis study).

Human psychology is enormously complex in the sense that there are many details to be learned that human beings have not yet understood. Once these details are brought to the light, we can easily grasp them. These details of human psychology must be brought to light, one by one, and then learned by all moral novices, so that they can act appropriately on them (or avoid acting inappropriately on them, as the case may be).

Nevertheless, it remains true that the process of moral development is a long one. Coming to know about implicit attitudes is but the first of many steps in unlearning or counteracting unjust implicit biases. But it is a necessary first step: individuals cannot even begin to undertake the process of overcoming unjust implicit attitudes themselves until they become aware of them. So it must be emphasized that the journey is only beginning when individuals begin to understand the implicit biases that they hold; but at the same time it must be emphasized that the journey *is* beginning and cannot begin without this crucial first step.

I end with one final worry. The findings of the IAT reveal patterns about the behavior of groups, not individuals. Virtue theory, by contrast, is focused on an individual's moral development. This disparity is significant: one cannot infer conclusions about a particular individual from findings about groups of individuals. But this point, while true, poses no threat to virtue theory or to the argument of this paper. First, virtue theory is important because it provides a clear framework in which individuals can be assessed for states that are beyond their

conscious choice and control. The virtue theorist has made this case regarding moral emotions and we have seen that this argument can now be extended to include implicit associations. So virtue theorists can offer a compelling argument for how we can morally assess individuals for their implicit attitudes, attitudes which are, in the moment they influence action, unchosen. Second, it must be acknowledged that some participants in the IAT have implicit attitudes that are more just, more virtuous, than other participants. This idea too can be easily explained by virtue theory: virtue theorists hold that some individuals are more expert in morality than others. So just as some individuals are expert cellists and others are novices, some individuals have excellent moral discernment and others are still developing this skill. Moral development is a process of acquiring a skill; the fact that some individuals exhibit little or no automatic association between black and good or between white and good may simply suggest that those individuals have cultivated the virtue of justice more than some of the other IAT participants. In short, the former are more virtuous than the latter. This idea coheres well with a virtue theory. Finally, one might hope that with more data about what distinguishes these different groups (those with just implicit associations from those with unjust ones), individuals might better understand what distinguishes these two groups and this, in turn, might empower those who love the good to move ourselves closer to the ideals of individual and social justice.

6. Aristotle argues that individuals cannot be held accountable for whether they feel emotions, but nonetheless can be held accountable for the way in which they feel them and the time at which they feel them and the way, over time, that they cultivate them. If individuals find themselves feeling and expressing too much anger, they can take actions to mitigate the

anger that they both feel and express. In particular, they can work to change the beliefs that give rise to the anger. And over time, this can begin to diminish the anger they feel in relevantly similar circumstances.

In the same way, individuals may not be accountable for whether they have implicit associations: indeed, unbeknownst to them, they may receive them from the cultures in which they live. Nevertheless individuals can hold themselves accountable for how they respond to the particular implicit associations that they have. If they find, as most of us do, that they associate white race with good and black race with bad, women with home and men with career, then they can take steps to mitigate these implicit associations. Implicit associations, from the perspective of the virtue theorist, are thus no different from any other moral particular: individuals can choose to be insensitive to them, or they can choose to learn about them, so that they can become more virtuous. Without a doubt, this is a long and arduous process. But becoming aware of unjust implicit associations is a necessary first step.

Both with emotions and with implicit associations, individuals can be subject to moral assessment for states that arise through their choices or for states that arise through the absence of choice. Furthermore, we can see the need for society to make its citizens aware of these implicit biases and implement programs to reduce and hopefully, one day not too long from now, eliminate these harmful associations. The more society is just the more its constituents can overcome this element of moral luck and come, themselves, to be just.ⁱ

Works Cited:

Aristotle, *The Nicomachean Ethics*, trans. W.D. Ross (1925) Oxford, UK: Oxford University Press.

Mahzarin R. Banaji, Kristi M. Lemm, and Siri J. Carpenter (2001) "The Social Unconscious." In A. Tesser & N. Schwartz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* Oxford, UK: Blackwell Publisher.

Miles F. Burnyeat (1980) "Aristotle on Learning to be Good." In *Essays on Aristotle's Ethics*, edited by Amelie O. Rorty, Berkeley and Los Angeles: University of California Press.

Nilanjana Dasgupta and Anthony G. Greenwald (2001) "On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals." *Journal of Personality and Social Psychology*, 81:5 pp. 800-814.

Alexander R. Green, MD, MPH, et al., (2007) "Implicit Bias Among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients," *Society of General Internal Medicine* (22: 1231-1238)

Rosalind Hursthouse (1999) *On Virtue Ethics*. Oxford: Oxford University Press.

<https://implicit.harvard.edu/>.

Thomas Nagel (1979) "Moral Luck," *Mortal Questions*. Cambridge, Mass: Cambridge University Press.

Alan Nelson, (2002) "Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care," Opening Statement from March 22, 2002, briefing of the Institute of Medicine, Washington DC. *Journal of the National Medical Association* 94: 8.

Brian A. Nosek, Mahzarin R. Banaji, and Anthony G. Greenwald, (2002) "Harvesting Implicit Group Attitudes and Beliefs From a Demonstration Web Site," *Group Dynamics: Theory, Research and Practice* 6: 1.

Laurie A. Rudman and Richard D. Ashmore, (2007) "Discrimination and the Implicit Association Test," *Group Processes and Intergroup Relations* 10: 3 pp. 359-372.

Nancy Sherman (1993) "The Role of Emotions in Aristotelian Virtue." In *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, Vol. IX (1993). Edited by John J. Cleary and William Wians. Lanham, MD: University Press of America, Inc. pp. 1-33.

Colin Tucker Smith and Brian A. Nosek, (2010) "Implicit Association Test." In I. B. Weiner & W. E. Craighead (Eds.), *Corsini's Encyclopedia of Psychology, 4th Edition*, Wiley Publisher.

Bernard Williams (1981) "Moral Luck," *Moral Luck*. Cambridge, UK: Cambridge University Press.

ⁱ I am indebted to Lauren Ashwell, Frank Chessa, Sarah Conly, David Cumiskey, Paul Stark, Charlotte Witt, and three anonymous reviewers at The Journal of Theoretical and Philosophical Psychology for wise and helpful comments on earlier drafts of this essay. I am grateful for their insight, which has done much to make this paper better.