Vetenskapsrådet

Ansökan

2017-03-21

Jönsson, Martin

Information om sökande

Namn: Martin Jönsson

Dr-examen: 2008-10-04

Födelsedatum: 19770614

Akademisk titel: Docent

Kön: Man

Arbetsgivare: Lunds universitet

Medelsförvaltare: Lunds universitet

Hemvist: Filosofiska institutionen 500011

Information om ansökan

Utlysningsnamn: Forskningsbidrag Stora utlysningen 2017 (Humaniora och samhällsvetenskap)

Bidragsform: Projektbidrag

Sökt inriktning: Fri

Ämnesområde utlysning: HS

Projekttitel (svenska): En kunskapsteoretisk undersökning av interventioner mot implicita fördomar

Projektstart: 2018-01-01

Projektslut: 2021-12-31

Sökt beredningsgrupp: HS-D

Klassificeringskod: 60301. Filosofi

Nyckelord: Implicita Fördomar, Epistemologi, Representationer

Sökta medel

År:	2018	2019	2020	2021
Belopp:	970 039	953 961	977 448	1 001 524

Beskrivande information

Projekttitel (svenska)*

En kunskapsteoretisk undersökning av interventioner mot implicita fördomar

Projekttitel (engelska)*

The Epistemology of Implicit Bias Intervention

Abstract (engelska)*

The project concerns *implicit biases:* automatic, unconscious associations that underlie prejudiced behaviour. More specifically it focuses on the methods – *implicit bias interventions* – we can use to overcome bias that causes us to *misrepresent* members of certain social groups (e.g. immigrants as dishonest, or women as unfit leaders).

Unreflectively, interventions that make *people* less biased (PI) might appear best; if people are biased, we should get them to stop! The project questions this idea, and thus a common conviction in the literature. It will

1) re-evaluate exactly what an intervention should correct in light of recent empirical developments

2) examine PI in light of several hitherto unrecognized problems:
- PI separates the discovery and correction of bias
- The accuracy of PI co-varies with the correctness of controversial psychological theses
- PI is likely to have unintended side-effects

3) systematically organize alternative implicit bias interventions in terms of their epistemological underpinnings and work out novel alternatives e.g. interventions that modify the results of the biased person's behaviour after the fact.

The project aims to significantly broaden research on implicit bias intervention and thus help combat racism and sexism. The project will be carried out over four years using standard philosophical methods; parts 1 and 2 will each take about 8 months; part 3 about 2 years; and the remaining time will be spent summarizing the results in a monograph.

Populärvetenskaplig beskrivning (svenska)*

Socialpsykologisk forskning från de senaste decennierna har visat att många av våra beslut påverkas av implicita fördomar—automatiska, undermedvetna tendenser att missgynna vissa sociala grupper. Speciellt oroande är att dessa fördomar hittas även hos t.ex. övertygade antirasister och feminister, i såväl privilegierade som i missgynnade grupper. Fördomarna genererar och vidmakthåller inte bara betydande orättvisor, de ger också oss skäl att tvivla på vad vi tror oss veta om människor i de missgynnade sociala grupperna. I ljuset av implicita fördomar, hur kan vi veta att vår uppskattning av en persons kompetens är korrekt, eller vad hon är benägen att göra eller hur troligt det är att hon har begått ett brott? Det här projektet handlar just om de implicita fördomar som får oss att representera varandra felaktigt (t.ex. invandrare som oärliga, kvinnor som dåliga ledare eller svarta som våldsamma).

I takt med att vår kunskap om implicita fördomar har ökat har forskare försökt utveckla metoder, vanligen benämnda som *interventioner*, för att komma till rätta med fördomarna. Dessa har varit av olika slag, t.ex. har man provat att exponera försökspersoner för bilder på medlemmar av missgynnade grupper som inte är i enlighet med de relevanta fördomarna, och man har utvecklat träningsprogram av olika slag där försökspersoner försöker arbeta bort fördomarna. Gemensamt för en väldigt stor del av den här forskningen är att den inriktat sig på *personliga interventioner*, interventioner som försöka göra *personer* mindre fördomsfulla. Det här projektet ifrågasätter den här intuitiva idén och därmed inriktningen på en stor del av den forskning som föreligger. Bakgrunden till projektet är dels att de personliga interventionerna empiriskt i väldigt hög utsträckning visat sig vara långsiktigt verkningslösa och dels att det finns principiella skäl som inte tidigare uppmärksammats att misstro personliga interventioner.

Projektet består av tre delar.

1) Projektet avser etablera, i ljuset av nya empiriska rön, exakt vilka förvrängningar som skall korrigeras av våra interventioner mot implicita fördomar.

2) Projektet kommer att kritiskt granska personliga interventioner utifrån ett antal problem som inte tidigare uppmärksammats:
- Att personliga interventioner genom sin natur separerar upptäckten av fördomar (att någon uppvisar ett fördomsfullt beteende) från hur dessa ska åtgärdas.
- Att personliga interventioners tillförlitlighet beror på huruvida kontroversiella psykologiska teser om deras natur är korrekta eller inte.
- Att fungerande personliga interventioner troligen kommer att ha oönskade sidoeffekter.

3) Projektet kommer att systematiskt organisera olika sorters interventioner i termer av deras epistemologiska status (t.ex. vilka antagande de vilar på och hur tillförlitligt de är) och utveckla nya sorters interventioner i mer detalj, t.ex. interventioner som korrigerar *resultatet* av ett fördomsfullt beteende (t.ex. genom att i efterhand uppdatera en fördomsfullt producerad rangordning), eller interventioner som manipulerar den fördomsfulle personens miljö för att undvika att hennes fördomar manifesteras.

Projektet strävar efter att bredda forskningen kring interventioner mot implicita fördomar. Genom att mer förutsättningslöst granska olika former av interventioner bidrar förhoppningsvis projektet till att identifiera verkningsfulla och välunderbyggda interventioner som kan fungera som instrument för att motverka den sexism och rasism, och annan form av diskriminering, som de implicita fördomarna ger upphov till.

Forskningsbeskrivning

Redogörelse för etiska överväganden*

Inga specifika etiska frågor uppkommer då projektet är av teoretisk karaktär och inte kommer att samla in persondata eller utföra experiment.

I projektet ingår hantering av persondata

Nej

I projektet ingår djurförsök

Nej

I projektet ingår humanförsök

Nej

Forskningsplan*

Se nästa sida för bilaga.

Purpose and aims

The project concerns *implicit biases*, understood as automatic, unconscious associations that underlie prejudiced behaviour. More specifically it focuses on the methods – *implicit bias interventions* – we can use to overcome the biases that cause us to systematically *misrepresent* members of certain social groups (e.g. misrepresent immigrants as dishonest, women as unfit leaders, men as poor caregivers, or black people as violent).

The project contains three interconnected parts.

1) It aims to re-evaluate to what extent our representations (e.g. beliefs, mental imagery, or rankings) of other people are distorted by implicit bias in the light of recent empirical developments (described in the next section).

2) It will critically examine the widespread idea – presupposed in much of the literature on implicit bias intervention – that a good way to come to terms with the distortions due to implicit bias is to make *people* less biased. A central aim of the project is to investigate how epistemologically well founded this idea is in light of a number of hitherto unrecognized problems (described in the project description).

3) It will systematically organize alternative implicit bias interventions in terms of their epistemological underpinnings and work out promising new alternatives in detail, among them *partially insulating interventions* and *post-hoc interventions*.

The project builds directly on earlier work of mine on implicit bias intervention (Jönsson and Sjödahl 2016; in preparation, and Jönsson submitted) and is in the same general vein as earlier work I have carried out on fallacious and biased reasoning (Jönsson and Hampton 2006, Jönsson and Assarsson 2013; 2016, Jönsson 2015, and Jönsson and Shogenji submitted).[1]

As the project progresses, articles corresponding to each part of the project will be published in international peer-reviewed journals. The majority of the final year of the project (the project will be carried out from 2018 – 2021) will be spent summarising the results of the project in a monograph intended for a main academic publisher. The monograph will give a much-needed overview of the epistemological advantages and disadvantages of different kinds of implicit bias interventions, and should thus help steer future research, policy and practice on how to overcome implicit bias.

Survey of the field

The idea that there are unconscious mental states that control our behaviour in undesirable ways is not new to psychology; Freud speculated at the beginning of 20th century that the unconscious was a repository for socially unacceptable ideas and desires that, although unknown to us, controlled various aspects of our behaviour. What was added to this idea with the inception of the research on implicit bias towards the end of the 20th century was a set of ideas on how to scientifically measure our unconscious mental states. Early seminal work in this regard is due to Greenwald, McGhee, and Schwartz (1998) who introduced the *Implicit Association Test* (IAT). Many others have suggested other tests (see e.g. Fazio 1995; Benaji and Hardin 1996; Nosek and Banaji 2001 and Payne et al., 2005) but the IAT has overwhelmingly been the most popular

[1] This is a revised version of the research plan I submitted to the Swedish Research Council last year. The main changes are as follows: 1) The plan no longer incorporates a sub-project to be carried out by a Ph.D.-student, since the intended Ph.D.-student (and the relevant sub-project) is now funded independently by Umeå University, 2) The plan no longer incorporates an empirical component (but will only use standard analytical philosophical methods) concerning implicit bias in ranking situations since this component is also be funded independently (by the research quota of 25% in my current position provided by the Faculties of Humanities and Theology, Lund University), 3) The plan has been adjusted in the light of new relevant research by myself (Jönsson and Sjödahl forthcoming; in progress, and Jönsson submitted) and others, in particular, an important paper by Forscher et al (submitted).

one (Greenwald et al's original article is cited over 8000 times according to Google-scholar) and will therefore be the focus of this review.

The IAT has many variations but centrally involve a simple categorization task with labels for social categories (e.g. the words "white"/"black" or "man"/"woman") and either *semantic* labels, i.e. labels that express a property or entity that the members of the social categories can have or be ("athletic", "stupid", "leader", "care-giver"), or *affective* labels, i.e. labels that express some sort of attitude ("good", "bad"). Participants are presented with stimuli (either words or pictures) that exemplify one of the aforementioned categories (e.g. pictures of white people and black people or things exemplifying "athletic" or "stupid"). Response times are then measured for categorization judgments that involve disjunctions of categories (e.g. how long it takes to categorize something as belonging to the disjunctive category "black or good" or "white or good") and response times (or a function of these) is taken to measure the implicit association between the disjoined concepts. Among many other things, the IAT has demonstrated e.g. greater association between pleasant words and white faces than between pleasant words and black faces (Nosek et al 2002), and greater association between "intelligence" and white faces than between "intelligence" and black faces (Amodio and Devine 2006).

Typically, the associations discovered by the IAT (and similar tests) are held to be implicit due to poor correlations between them and verbal reports of participants' beliefs and attitudes. Moreover, the associations are held to be important to the extent that the implicit measures are correlated with actual discriminatory behaviour. A recent comment on two large-scale meta-analyses on the predictive validity of the IAT maintained that

> "[A]mong the settings in which IAT measures can be used to predict discrimination are personnel decisions (hiring, performance evaluation, salary, promotion), law enforcement decisions (stops and searches of drivers, pedestrians, or travellers), criminal justice decisions (jury and bench verdicts, sentencing, bail setting, parole, inmate discipline), educational decisions (admissions, grading, disciplinary actions, suspensions), and health-care decisions (triage, treatment authorization, prescription)." (Greenwald et al 2015, p. 557).

It should be mentioned though, that both the relevant meta-analyses found the average predictive validity correlation of IAT to be quite low ($r = .148$ and $r = .236$ respectively)

Here are a few specific examples of discriminatory behaviour that has been linked to implicit bias in the literature (cf. Saul 2012): 1) CV-studies reveal that the same CV is considered much better when it has a typically white rather than a typically black name, a typically Swedish rather than typically Arab name, or a typically male rather than typically female name (Bertrand and Mullainathan 2004; Rooth 2007; Moss-Racusin et al. 2012; Steinpreis et al. 1999), 2) Studies of so-called 'shooter bias' reveal that the very same ambiguous object is far more likely to be perceived as a gun when held e.g. by a young black man or a man who appears to be muslim than when held by a young white man. (Correll et. al. 2002, 2007; Greenwald et al. 2003; Payne, 2001; Unkelbach et al. 2008), and 3) Studies of so called 'prestige bias' reveal that top psychology journals to a large extent reject previously published papers when these are resubmitted with false names and non-prestigious affiliations (Peters and Ceci 1982).

During the last decade, implicit bias has increasingly attracted the attention of philosophers. Within epistemology alone, a number of questions have so far been discussed; whether there is a connection between being prejudiced and being epistemically blameworthy (Begby 2013), what the epistemic costs of living in a prejudiced society are (Gendler 2011; Egan 2011; Madva forthcoming), and how the prejudiced discounting of testimony leads to epistemic injustice (Fricker 2007; Hookway 2010; Anderson 2012).

However, the central work for the purposes of this project is Saul (2012) who argues that implicit bias (understood as automatic tendencies to associate certain traits with members of particular social groups that are unavailable to inspection and rational evaluation) gives rise to

something akin to a new form of scepticism, what she calls 'bias-related doubt', and that we have very good reason to believe that we cannot properly trust our knowledge-seeking faculties

Saul goes on to discuss what we should do to improve our epistemological situation. Although she maintains that to fully combat the influence of implicit biases, we need to re-shape our social world, she endorses a number of small-scale suggestions from the psychology of implicit bias intervention, e.g. 1) spend time thinking about counter-stereotypical exemplars (members of stereo-typed groups who don't fit the group stereotypes) (Blair 2002, Kang and Banaji 2006), 2) carefully form implementation intentions like 'when I see a black face I will think 'safe'' (Stewart and Payne 2008), or 3) or spend a few hours engaging in Kawakami's negation training, in which one practices 'strongly negating stereotypes' (Kawakami et al. 2000; Johnson 2009).

Saul's suggestions are in line with the psychological literature on implicit bias interventions (cf. Lai et al 2014; Forscher et al 2016) in the sense that they heavily focus on *personal interventions*, interventions that try to change the mind (/brain) of the biased person in order to overcome bias.[2] A recurring pattern in this literature however is the difficulty of successfully making people less biased. Becoming aware that one is biased does not translate directly into how to improve one's behaviour, and bias 'remains stubbornly immune to individual efforts to wish it away' (Hardin and Banaji, 2013). Some studies show that it can even be directly counterproductive to explicitly ask people to improve (Legault et al, 2011). Moreover, even when successful, interventions tend to reduce very specific biases. For instance, Amodio and Lieberman (2009) argue that semantic and affectual biases need different interventions (see also Forbes and Schmader 2010).

Although there is some support for the effectiveness of the interventions that Saul mentions, recent large-scale investigations of personal interventions have not offered very much support. Lai et al. (2014) and Lai et al. (2016) demonstrated (in a study with over 15 000 participants) that out of 17 interventions tested, only 9 had significant effects on IAT-scores, *and none of these had long term effects*.[3] Moreover, Forscher (submitted) showed that even when interventions successfully lower IAT-scores, they do not significantly diminish the biased behaviour that the IAT-scores are correlated with.

I think that these findings have a number of important consequences for the epistemological question that Saul (2012) was interested in: 1) In order to successfully overcome implicit bias very specific strains of bias needs to be targeted, 2) Since personal interventions have such a poor track-record, we should take a step back and consider the reasons we have to think that such interventions will really improve our epistemic situation, and 3) we should see if we can develop new kinds of interventions with more robust epistemological underpinnings.

This project will investigate how our epistemic situation can be improved in the face of implicit bias in line with these points. It will focus specifically on the biases that cause us to misrepresent members of certain social groups (i.e. immigrants as dishonest, women as unfit leaders, men as poor caregivers, or black people as violent). Moreover, it will consider non-personal interventions that attempt e.g. to modify the environment of the biased person (*environmental interventions*), or modify the results of the biased persons behaviour after the fact (*post-hoc interventions*) rather than modify the person herself. Some environmental interventions have been discussed in the literature. For instance Saul (2012) discusses anonymization as a way to avoid bias. This is a form of *insulating* environmental intervention where the agent's environment is modified so that her bias won't be manifested (since she is ignorant of the categories that triggers it). Jönsson (submitted) have considered a partially insulating intervention (described below). Other environmental interventions such as facilitating a diverse working environment have also been discussed (cf. Jolls and Sunstein 2006). Post-hoc

[2] Personal interventions need not be "personal" in the sense of the philosophical distinction between "personal" and "subpersonal".

[3] No effect of the intervention could be detected 2-4 days after the intervention had occurred.

interventions have not been discussed at all until Jönsson and Sjödahl (2016) and will be described below (in Part 3).

Project description

Part 1: The extent to which implicit bias distorts our representations of other people

The first part of the project aims to establish to what extent our representations of each other are distorted by implicit bias. By focusing on these *particular* epistemologically detrimental effects, we are better prepared to identify ways to overcome them (in parts 2 and 3) than if we focus on a more inclusive class of effects, which might have separate causes.

To determine the extent to which implicit bias distorts our representations we need to distinguish two things; *the bias itself*, and *its manifest consequences*. Both might be representational. If biases are general beliefs (such as the belief that men are better leaders than women), they are themselves representational. In addition, they might give rise to other representations such as utterances (e.g. the utterance "John will be a better leader than Wanda"), or rankings (e.g. an ordering of a set of candidates for a job where men are inappropriately generally ranked higher than women). Even if biases are not beliefs but some looser set of associations (e.g. *aliefs*, c.f. Gendler 2008) they might still be representational, and even if they are not, they might still give rise to representational consequences.

Much of the philosophical literature on implicit bias explicitly maintains that implicit bias makes us misrepresent members of disadvantaged groups very generally, and that we thus face a new form of scepticism (c.f. Saul 2012 for a clear example). But there are a number of reasons in the literature that suggest that misrepresentation due to implicit bias, although it occurs, is less frequent than has been supposed.

First, as we have seen above, the implicit associations studied in the literature are usually subdivided into affective and semantic associations. Both of these might lead to discriminatory behaviour, but only the latter associations can *themselves* be distortions. Affective associations are not distortions, since they do not purport to represent anything at all. To have a feeling of disgust associated with immigrants is not in itself to represent them as having a certain property they do not have, even though it is problematic for other reasons. So, evidence of affective associations is not in itself any evidence for distortions.

Second, the semantic associations, by definition, involve representations (e.g. "black" and "violent"). However, they can only be said to incorporate a misrepresentation or distortion if the associations correspond to beliefs, or something belief-like. The co-occurrence of two representations is not the same thing as predicating the content of one on the content of the other. So whether evidence of semantic associations is also evidence of distortion depend on whether they correspond to beliefs, which is controversial (see e.g. Gendler 2008; 2011; Egan 2011; Mandelbaum 2014).

Third, even if it turns out that semantic associations generally are beliefs, they only count towards the distortions due to implicit bias *if they really are implicit*. In many cases however, what appeared initially to be an *implicit* bias turned out not to be implicit. Following Bargh (1994), Gawronoski et al., (2006) distinguish between *content-awareness* (being aware that you believe, e.g. that immigrants are dishonest), *source-awareness* (being aware why you believe, e.g. that immigrants are dishonest), and *impact-awareness* (being aware how your behaviour is influenced by the fact that you believe that immigrants are dishonest). Although we do not generally have much source awareness or impact awareness of implicit biases, there is evidence that the correlation between indirect measures of bias and self-reports can be increased by increasing participant's motivation for being honest, or decreasing the amount of deliberation that precedes a self-report (Gawronoski et al. 2006). This suggests that people have a higher degree of content-awareness of implicit biases than has previously been assumed, and thus that they are not really implicit (in the sense of "unconscious").

Fourth (and finally), in her discussion of bias-related-doubt Saul (2012) doesn't distinguish between a) the claim that implicit bias makes us believe erroneous things about members of certain social groups, b) the claim that implicit bias lowers the credence we attribute to statements, texts, or testimony more generally, when produced by members of relevant social groups. Although both claims are important, it should be noted that among the support she cites for her position, only CV-studies and the studies of 'shooter bias' support the first claim (studies of prestige bias support the second), and thus that the support for the misrepresentation *of other people* due to implicit bias is more limited than it first appears in her discussion.

We can conclude that there are a number of factors that give us good reasons to reconsider the empirical data on implicit bias to pinpoint the circumstances under which implicit bias gives rise to misrepresentation.

Part 2: Problems with personal approaches to overcome implicit bias

The second part of the project aims to evaluate (in light, partly, of the results of the first part of the project) the prevalent idea that a reasonable way to overcome implicit bias is to make *people* less biased, i.e. to make use of personal interventions. Unreflectively, personal interventions might strike one as obviously the best way to come to terms with implicit bias; if people are biased, we should try to get them to stop! However, this approach has both empirical and theoretical problems. Before we get to these we can consider a point that somewhat undermines the presumption that implicit bias should be handled by personal interventions: one reason for believing that personal interventions are preferable to non-personal alternatives might be the conviction that the personal interventions are ways to combat something more general than the biased person's behaviour on some particular task; it is clearly more desirable to make a person less biased, than to just change her behaviour on a particular occasion. However, as was mentioned above there is accumulating evidence that personal bias interventions, when successful, do not generalize very much (cf. Amodio and Lieberman 2009; Forbes and Schmader 2010; Madva and Brownstein forthcoming).

The empirical problem

First, as the survey of the field made clear, recent large scale studies of personal interventions representing "state-of-the-art knowledge about implicit attitude change" (Lai et al 2014; Lai et al 2016) have called into question their effectiveness; out of 17 interventions tested only 9 had significant effects on iat-scores, *and none of these had long term effects*. These findings gives us strong reasons to be critical of personal interventions.

The incommensurable quantification problem

Second, what we are epistemologically interested in when we are dealing with a misrepresentation is *a more accurate* representation, not just a different representation. But discovering that a person exhibits a biased behaviour to some degree *x* give us no indication of how to calibrate a personal intervention as a function of the degree of biased behaviour to improve her representations. For instance, assume that a person generally gives men twice as many points as women on an evaluation. What does this mean in terms of how much that person must go through implementation training, or how much time that person should spend with counterstereotypical exemplars, to achieve accurate representations? For personal interventions, the discovery and correction of bias are divorced; a measure of misrepresentation is not translatable into the quantities of personal interventions. Post-hoc interventions that correct biased behaviour after the fact are at a clear advantage here, since these are solutions framed in the same terms as the problems they adress.

The mental posits problem

Personal interventions build on theorizing about the mental and neuronal underpinnings of implicit bias. The appropriateness of these interventions are thus contingent on the correctness

of this theorizing. Mutatis mutandis, our conviction in the veracity of the representations of the biased person after an intervention, is contingent on the correctness of the assumptions about the underlying reasons for the bias. The disagreement in the literature about several aspects of the underlying nature of implicit bias – e.g. whether or not the bias is a belief or not (cf. Gendler 2008; 2011; Egan 2011; Mandelbaum 2014), or whether semantic or affectual association correspond to two constructs (cf. Glaser, 1999; Correll et al., 2007; Amodio & Devine, 2006; Forbes & Schmader, 2010) or just one (cf. Greenwald et al., 2002; Madva and Brownstein forthcoming) – thus gives us a general reason to doubt the that personal interventions offer epistemological improvements. When theorizing is inconclusive, the utility of methods that depend on that theorizing is indeterminate.

The problem of unwanted consequences

It is a mistake to think that implicit bias is somehow external to normal (unbiased) capacities. Implicit bias intervention is thus unlike surgically removing a tumour that hasn't spread to adjacent tissue. It follows that any personal interventions that have effect, is likely to have unintended additional consequences (e.g. unintended changes to our conceptual system). Whether or not a personal intervention is epistemologically beneficial then, is not something that can be established on the basis of improvment on an implicit bias test, since it doesn't measure changes that are independent of the bias, but potentially epistemologically detrimental.

Part 3: A taxonomy of alternative approaches to overcome bias and new interventions

The third part of the project aims to systematically review and taxonomically organize possible general ways to overcome bias in terms of their epistemological status, e.g. the plausibility of the assumptions they rely on, and how likely they are to improve the veracity of our representations.

Although some attempts have already been made in the literature to organize ways to overcome bias, these have not been very general (e.g. Jolls and & Sunstein 2006, Lai et al 2014), and they have never been attempts to divide interventions at the epistemological joints.

As a preliminary rough division we can distinguish between the already mentioned *personal*, *environmental* and *post-hoc interventions*. Among the environmental interventions we can distinguish at least between *insulating* and *non-insulating* interventions. Insulating interventions are those that try to prevent people from seeing or doing certain things in order to prevent biases from triggering, and non-insulating environmental interventions are those that try to modify a persons environment in order to moderate the expression of her bias. Where anonymization is possible, it is the most powerful among the insulating interventions since it completely shuts off the bias by hiding relevant group memberships. Less powerful, *partially insulating interventions* are also possible. A number of studies have suggested that implicit bias is especially prevalent in the aggregation of merits in hiring and admission processes (e.g. Dovidio och Gaertner 2000; Hodson, Davidio och Gaertner 2002; Uhlman och Cohen 2005). In light of these findings, Jönsson (submitted) compared different methods to aggregate merits when hiring Ph.D.-students (according to the rules and practices at Lund University). Each of the methods were evaluated in terms of its 1) plausibility, 2) it's underlying assumptions (in particular its measurement assumptions, such as whether it presupposed an ordinal or an interval scale) and 3) its potential to insulate the person(s) doing the aggregation from being influenced by bias.

The project will investigate partially insulating interventions like this one in more general terms. It seems, for instance, that partly insulating interventions face a variant of the mental posits problem in that they presupposes knowledge about the circumstances that gives rise to biased behaviour. Moreover, the conclusions of Jönsson (submitted) need to be generalized beyond the particular situation it was concerned with (hiring Ph.D.-students according to the rules and practices at Lund University) to be relevant in a wider context. This will also involve bringing to bear a wider range of methods from the field of Multi-Criteria Analysis (cf. Ishizaka and Nemery 2013) on implicit bias intervention.

Another important focus of this part of the project is on post-hoc interventions. In a recent article (Jönsson and Sjödahl 2016) we argued that in ranking situations (for grants, jobs or promotions), we have access to a non-personal form of intervention, that has not been considered in the literature; on this view rather than trying to change the person(s) producing rankings, we could change the ranking itself. By using an evaluator's history of past rankings, logic and statistics can be used to reliably improve a new ranking without thereby first having to change the evaluator. Our argument in the article depended (among other things) on an equal qualifications assumption according to which the property *p* that the rankings are supposed to track is evenly distributed in the populations corresponding to all relevant social groups. For instance, if we are only interested in men and women, we assume that *p* occurs with the same frequency and magnitude in the population of all men and the population of all women. Given this assumption, and access to a history of past rankings we argued that a set of functions can be identified such that wherever these converged, the original ranking could be updated accordingly and would, if updated, reliably be more correct than it was initially. This form of intervention avoids all four problems described in part 2 above; 1) the empirical problem is not applicable, 2) the measure used to discover the bias is directly used to correct the bias, 3) we need not posit anything about the workings of the biased persons mind, and 4) the intervention only affects exactly that we want to change.

Post-hoc interventions are not without problems however and some of these will be addressed in the project. In particular, the form of post-hoc intervention suggested by Jönsson and Sjödahl (2016) depended on a number of assumptions (such as the equal qualifications assumption) and it needs to investigated whether or not these could be weakened. Moreover, the basic idea of a post-hoc intervention is to overrule an evaluator's initial estimates. Such overruling can lead evaluators to trying to 'game' the corrective measure to push through what they believe to be the correct ranking. The robustness of the intervention in the face of such behaviour will be investigated. A final problem for post-hoc interventions is that they are not applicable to "ballistic" implicit bias where there is no possibility to correct for a misrepresentation before an undesirable consequence (most obviously illustrated with shooter bias). The generality of post-hoc interventions thus needs to be worked out.

The overall division in terms of personal, environmental and post-hoc interventions should be seen as tentative, and is likely to change during the course of project. Another aspect to the project, that is also difficult to specify in advance, is which additional kinds of interventions that will be explored. One kind of intervention that hasn't received much attention in the literature is the form of environmental interventions that occur when people make small changes to their own behaviour – not to become less biased themselves – but to shape each other's environments (e.g. maintaining that a person from a stigmatized group is more competent than another person, whenever one judges the two persons to be equally competent). It is interesting to explore under what circumstances such micro-changes can generate positive macro changes.

Time Table

The project will carried out at the Department of Philosophy, Lund University over a period of four years, from 1 Jan 2018 to 31 Dec 2021, 75% of fulltime. Each of the two first parts are estimated to take about 8 months, while the third, more exploratory, part is estimated to take slightly less than 2 years. As the project progresses, articles corresponding to each part of the project will be published in international peer-reviewed journals. The majority of the remaining time (/the fourth year) will be spent summarizing the results in a monograph.

Significance

The project will significantly broaden the research on implicit bias intervention. It is novel in several ways: 1) implicit bias interventions have not themselves previously been subject to epistemological analysis, 2) the (non-empirical) problems posed for personal interventions have not previously been articulated or discussed, 3) post-hoc interventions have not previously been

discussed. Moreover, through the project's unique vantage point, it will likely be able to identify additional kinds of implicit interventions as it progresses. The monograph that will summarize the results of the project will be a much-needed comprehensive review of the assumptions underlying different kinds of bias interventions and their respective chances of improving the veracity of our representations. It will thereby give a balanced – currently absent in the literature – view on how to best further research implicit bias intervention.

The results of the project might also have wider philosophical significance and contribute towards central topics in epistemology (e.g. the possibility of self-knowledge, and first person authority) in philosophy of mind (e.g. how to think about the positing of unconscious mental states) and in ethics (e.g. moral responsibility in the face of implicit bias).

The project will also have practical significance in that it investigates how we can come to terms with a particular form of prejudiced behaviour. It can thus inform policymaking and thereby help combat racism and sexism. Moreover, post-hoc interventions are related to the practice of affirmative action by the use of e.g. gender-quotas. But whereas affirmative action typically works by overriding the perceived competence of candidates *not from disadvantageous groups* in a very blunt manner, post-hoc interventions modify e.g. rankings only to the extent that they can correct for a detected bias. This means that the results of this project could help improve societal reform by replacing the use of a very blunt instrument (quotas or quota-like mechanism) with a more fine tuned one (post-hoc interventions).

Preliminary results

As has already been described, the project builds on earlier philosophical work of mine on implicit bias intervention (Jönsson and Sjödahl 2016; Jönsson submitted). An embryo of Jönsson and Sjödahl (2016) was presented by Sjödahl at the conference 'Why are there so few women in philosophy?' held at Stockholm University, April 17th-18th 2015.

In addition, an empirical investigation (Jönsson and Sjödahl in preparation) of relevance to the project is already underway. The investigation attempts to isolate the conditions under which implicit bias influences the ranking of candidates for fictitious jobs. In a few preliminary experiments we asked participants (n=100 for each experiment) to rate, on a scale from 1 to 8, three sets of CV's corresponding to applicants for a job either as a fire-fighter, pharmacist or kindergarten teacher. Participants were assigned to one of two conditions and depending on their assignment they saw a particular CV with either a male or a female name. Mean ratings for each CV were compared across groups in a between-subjects design to detect bias. A post-test debriefing measured the participants' explicit attitudes towards men and women in the three professions in order to exclude the portion of the data pertaining to explicit bias.

The first experiment, which failed to reveal any bias, was subsequently used as baseline to which the effect of different manipulations could be compared. One such manipulation was the addition of a picture (from the Chicago Face Database, see Ma, Correll, and Wittenbrink, 2015) to each CV in order to make the gender of each candidate more salient. This experiment generated a number of instances of significant bias. Interestingly, in each case, the pair of pictures were not themselves sufficient to induce bias but did so only via an interaction with a particular CV. The general pattern seems to be that only "ambiguous" CVs (i.e. those where one of two kinds of merits e.g. "education" and "professional experience" was relatively strong and the other relatively weak) generated bias-effects (this is consistent with findings in Hodson, Dovidio, & Gaertner, 2002). Additional experiments are planed that will test the boundary conditions of the biased ranking.

The empirical investigation will support the project described here in at least two ways. First, the results from the experiments will provide data of biased ranking which can be used to validate post-hoc interventions. Second, by conducting controlled experiments where the effects of manipulating particular variables can be tested, data is generated that are relevant for the development of partially insulating interventions.

References: (sources separated by ❖ to conform to the required number of pages)

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behaviour. *Journal of Personality and Social Psychology*, 91(4), 652. ❖ Anderson, E., 2012, "Epistemic justice as a virtue of social institutions", *Social Epistemology*, 26(2): 163–173. ❖ Bargh, J., 1994, "The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition", in *Handbook of social cognition* (2nd ed.), R. Wyer, Jr. & T. Srull (eds.), Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., pp 1–40. ❖ Begby, E., 2013, "The Epistemology of Prejudice", *Thought*, 2(2): 90–99. ❖ Banaji, M. & C. Hardin, 1996, "Automatic stereotyping", *Psychological Science*, 7(3): 136–141. ❖ Bertrand, M., and Mullainathan, S. 2004. Are Emily and Greg more employable than Lakisha and Jamal? *American Economic Review*, 94, 991–1013. ❖ Blair, I. 2002. The Malleability of Automatic Stereotypes and Prejudice. *Personality and Social Psychology Review*, 3: 242-261. ❖ Correll, J., Park, B., Judd, C., & Wittenbrink, B. 2002. The police officer's dilemma. *Journal of Personality and Social Psychology*, 83, 1314–1329. ❖ Correll, J., Park, B., Judd, C., Wittenbrink, B., Sadler, M. S., & Keesee, T. 2007. Across the thin blue line. *Journal of Personality and Social Psychology*, 92, 1006–1023. ❖ Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2007). The influence of stereotypes on decisions to shoot. *European Journal of Social Psychology*,37(6), 1102-1117. ❖ Dovidio, J. F., & Gaertner, S. L., (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, 11, 319-323. ❖ Egan, A., 2011. "Comments on Gendler's 'The epistemic costs of implicit bias,'", *Philosophical Studies*, 156: 65–79. ❖ Fazio, R., 1995, "Attitudes as object-evaluation associations", in Attitude strength: Antecedents and consequences (*Ohio State University series on attitudes and persuasion*, Vol. 4), R. Petty & J. Krosnick (eds.), Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., pp. 247–282. ❖ Forbes, C. E., & Schmader, T. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive capacity under stereotype threat. *Journal of Personality and Social Psychology*, 99(5), 740. ❖ Forscher, P.S., Lai, C.K., Axt, J.R., Ebersole, C.R., Herman, M., Devine, P., and Nosek, B. A., submitted, A Meta-Analysis of Change in Implicit Bias, manuscript under review accessed via https://www.researchgate.net/publication/308926636_A_Meta-Analysis_of_Change_in _Implicit _ Bias, Accessed 2017-03-20 ❖ Fricker, M., 2007, *Epistemic Injustice: Power & the Ethics of Knowing*, Oxford: Oxford University Press. ❖ Gawronski, B., W. Hofmann, & C. Wilbur, 2006, "Are implicit attitudes unconscious?", *Consciousness and Cognition*, 15: 485–499. ❖ Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. ❖ Greenwald, A. G., Oakes, M. A. and Hoffman, H. 2003. Targets of discrimination: Effects of race on responses to weapons holders. *Journal of Experimental Social Psychology*, 39, 399–405. ❖ Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological review*, 109(1), 3. ❖ Greenwald, Banaji and Nosek (2015) Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects. *Journal of Personality and Social Psychology* Vol. 108, No. 4, 553–561 ❖ Gendler, T., 2008, "Alief and belief", *The Journal of Philosophy*, 105(10): 634–663. ❖ Gendler, T., 2011, "On the epistemic costs of implicit bias", *Philosophical Studies*, 156: 33–63. ❖ Glaser, J. C. (1999). *The relation between stereotyping and prejudice*. Doctoral dissertation, Harvard University. ❖ Hardin, C. D., & Banaji, M. R. (2013). The nature of implicit prejudice: Implications for personal and public policy. In E. Shafir (Ed.), *The behavioral foundations of policy*. ❖ Hookway, C., 2010, "Some Varieties of Epistemic Injustice: Response to Fricker", *Episteme*, 7(2): 151–163. ❖ Hodson, G., Dovidio, J.F., & Gaertner, S.L. (2002) Processes in racial discrimination: Differential weighting of conflicting information. *Personality and Social Psychology Bulletin*, 28, 460–471 ❖ Ishizaka and Nemery 2013 *Multi-Criteria Decision Analysis* John Wiley & Sons Ltd, Chichester, UK ❖ Johnson, I. R. 2009. *Just say 'No' (and mean it): Meaningful negation as a tool to modify automatic racial prejudice*. Doctoral dissertation, Ohio State University. ❖ Jolls, Christine and Sunstein, Cass R., "The Law of Implicit Bias" (2006). Faculty Scholarship Series. Paper 1824. http://digitalcommons.law.yale. edu/ fss_papers/1824 ❖ Jönsson, M. L. (2015) Overextension in

Verb-Conjunctions. *Journal of Experimental Psychology: Learning, Memory and Cognition* 41(6), 1917-1922. ❖ Jönsson, M. L. (Submitted) Allt sammantaget den bästa kandidaten. Manuscript under review, *Högre Utbildning* ❖ Jönsson, M. L. & Assarsson, E. (2013). Shogenji's measure of justification and the inverse conjunction fallacy. *Synthese* 190(15), 3075-3085. ❖ Jönsson, M. L. & Assarsson, E. (2015). A Problem for Confirmation Theoretic Accounts of the Conjunction Fallacy. *Philosophical Studies* ❖ Jönsson, M., and Sjödahl, J. (2016). Increasing the veracity of implicitly biased rankings. *Episteme*,1-19. https://doi.org/10.1017/epi.2016.34 ❖ Jönsson, M. L & Sjödahl, J. (in preparation). "Implicit Bias in the Rating and Ranking of Resumes", manuscript ❖ Jönsson, M. L. & Hampton, J. A. (2006). The Inverse Conjunction Fallacy. *Journal of Memory and Language*, 33(5), 317-334. ❖ Jönsson, M. L. & Shogenji, T. (submitted). A Unified Account of the Conjunction Fallacy by Coherence. Manuscript under review, *Synthese* ❖ Kang, J. and Banaji, M. 2006. Fair Measures: A Behavioral Realist Revision of 'Affirmative action'. *California Law Review* 94: 1063-1118. ❖ Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S. and Russin, A. 2000. Just say no (to stereotyping). *Journal of Personality and Social Psychology* 78 , 871–888. ❖ Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., Sartori, G., Dial, C. M., Sriram, N., Banaji, M. R., & Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, 143, 1765-1785. ❖ Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., Calanchini, J., Xiao, Y. J., Pedram, C., Marshburn, C. K., Simon, S., Blanchar, J. C., Joy-Gaba, J. A., Conway, J., Redford, L., Klein, R. A., Roussos, G., Schellhaas, F. M. H., Burns, M., Hu, X., McLean, M. C., Axt, J. R., Asgari, S., Schmidt, K., Rubinstein, R., Marini, M., Rubichi, S., Shin, J. L., & Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*. ❖ Legault, L., Gutsell, J., and Inzlicht, M. (2011). Ironic Effects of Antiprejudice Messages. *Psychological Science* 22(12) 1472–1477. ❖ Ma, Correll, and Wittenbrink, 2015, The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data, Behavioral Research Methods. 47:1122–1135 ❖ Madva, A., forthcoming, Biased against Debiasing, *Ergo*. ❖ Madva, A. & Brownstein (Forthcoming), "Stereotypes, Prejudice, and the Taxonomy of the Implicit Social Mind*, Noûs.* ❖Mandelbaum, E., 2014, "Attitude, Association, and Inference: On the Propositional Structure of Implicit Bias", *Noûs*. doi:10.1111/nous.12089 ❖ Moss-Racusin, C., Dovidio, J., Brescoll, V., Graham, M., Handelsman, J. 2012. Science Faculty's Subtle Gender Biases Favor Male Students. *PNAS* 109 (41) 16395-16396. ❖ Nosek, B. & M. Banaji, 2001, "The go/no-go association task", *Social Cognition*, 19(6): 625–666 ❖ Nosek, B. A.; Banaji, M. R.; Greenwald, A. G. (2002). "Harvesting implicit group attitudes and beliefs from a demonstration website". *Group Dynamics*. 6 (1): 101–115. ❖Payne, B. K. 2001. Prejudice and perception processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81, 181–192. ❖ Payne, B., C.M. Cheng, O. Govorun, & B. Stewart, 2005, "An inkblot for attitudes: Affect misattribution as implicit measurement", *Journal of Personality and Social Psychology*, 89: 277–293. ❖ Peters, Douglas P., and Stephen J. Ceci. 1982. Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences* 5:187–255. ❖ Rooth, D. 2007. Implicit discrimination in hiring: Real world evidence (IZA Discussion Paper No. 2764). Bonn, Germany: Forschungsinstitut zur Zukunft der Arbeit. ❖ Saul, J. (2012). Scepticism and Implicit Bias. *Disputatio* 5: 37, 243-263. ❖ Steinpreis, R., Anders, K., and 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*, 41: 7/8, 509–528. ❖ Stewart, B. D., and Payne, B. K. 2008. Bringing Automatic Stereotyping under Control: Implementation Intentions as Efficient Means of Thought Control. *Personality and Social Psychology Bulletin*, 34, 1332-1345. ❖ Uhlmann, E.L., & Cohen, G.L. (2005). Constructed criteria. *Psychological Science*, 16, 474-480. ❖ Unkelbach, C., Forgas, J. and 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* 44: 5, 1409- 1413.

Jag är projektledare för ett pågående fritt projektbidrag inom området Humaniora och samhällsvetenskap för vilket Vetenskapsrådets utbetalning av medel pågår t o m 2017

Budget och forskningsresurser

Aktivitetsgrad i projektet*

	Roll i projektet	Namn	Procent av heltid
1	Projektledare	Martin Jönsson	75%

Löner inklusive sociala avgifter

	Roll i projektet	Namn	Procent av lönen
1	Projektledare	Martin Jönsson	75%
Totalt			0

	2018	2019	2020	2021	Totalt
1	634 669	650 536	666 799	683 469	2 635 473
Totalt	634 669	650 536	666 799	683 469	2 635 473

Lokaler

	Typ av lokal	2018	2019	2020	2021	Totalt
1	Andel av värdinstitutionens lokalkostnader	45 718	44 960	46 067	47 202	183 947
Totalt		45 718	44 960	46 067	47 202	183 947

Driftskostnader

	Driftskostnader	Beskrivning	2018	2019	2020	2021	Totalt
1	Dator	Dator	27 000				27 000
2	Konferensresor	Konferensresor	10 000	10 000	10 000	10 000	40 000
Totalt			37 000	10 000	10 000	10 000	67 000

Avskrivningar utrustning

Avskrivning	Beskrivning	2018	2019	2020	2021
Ingen information ifylld					

Total budget*

	Specificerade kostnader	2018	2019	2020	2021	Totalt, sökt
1	Löner inkl. sociala avgifter	634 669	650 536	666 799	683 469	2 635 473
2	Driftskostnader	37 000	10 000	10 000	10 000	67 000
3	Avskrivningar utrustning					0
4	Lokaler	45 718	44 960	46 067	47 202	183 947
5	Delsumma	717 387	705 496	722 866	740 671	2 886 420
6	Indirekta kostnader	252 652	248 465	254 582	260 853	1 016 552
7	Total projektkostnad	970 039	953 961	977 448	1 001 524	3 902 972

	Annan kostnad	Total kostnad
1		2 635 473
2		67 000
3		0
4		183 947
5	0	2 886 420
6		1 016 552
7	0	3 902 972

Motivering av sökt budget*

Kostnaderna avser lönekostnader för mig själv på 75% av heltid inklusive LKP (50,04%) samt en engångskostnad för en dator samt årliga kostnader för konferensresor (inklusive indirekta kostnader). Indirekta kostnader motsvarar totalt 44,42% (varav 6,81% är lokalkostnader som angetts separat på rad 4)

Annan finansiering för detta projekt

Finansiär	Sökande/projektledare	Typ av bidrag	Status	Dnr eller motsv.
2018	2019	2020	2021	
Ingen information ifylld				

Publikationer

Publikationer (pdf)*

Se nästa sida för bilaga.

List of Publications

1. Peer-Reviewed Original Articles

2016 **Jönsson, M. L.** & Sjödahl, J. (2016) Increasing the Veracity of Implicitly Biased Rankings. *Episteme* https://doi.org/10.1017/epi.2016.34 ISSN: 1742-3600*

Jönsson, M. L. (2016) Interpersonal Sameness of Meaning for Inferential Role Semantics. *Journal of Philosophical Logic* doi:10.1007/s10992-016-9400-3 ISSN: 0022-3611

Jönsson, M. L. & Assarsson, E. (2016) A Problem for Confirmation Theoretic Accounts of the Conjunction Fallacy. *Philosophical Studies* 173(2), 437-449. ISSN: 0031-8116*

2015 **Jönsson, M. L.**, Hahn, U. and Olsson, E. J. (2015) The Kind of Group You Want to Belong to: the Effect of Group Structure on Group Performance. *Cognition* 142, 191-204. ISSN: 0010-0277

Jönsson, M. L. (2015) Overextension in Verb-Conjunctions. *Journal of Experimental Psychology: Learning, Memory and Cognition* 41(6), 1917-1922. ISSN: 0278-7393*

Jönsson, M. L. (2015) Linguistic Convergence in Verbs for Belief-Forming Processes. *Philosopical Psychology* 28(1-2), 114-138. ISSN: 0951-5089

2014 **Jönsson, M. L.** (2014) Semantic Holism and Language Learning. *Journal of Philosophical Logic* 43(4), 725-759. ISSN: 0022-3611

2013 **Jönsson, M. L.** & Assarsson, E. (2013) Shogenji's measure of justification and the inverse conjunction fallacy. *Synthese* 190(15), 3075-3085. ISSN: 0039-7857*

Jönsson, M. L. (2013) An Empirically Grounded Solution to the Generality Problem. *Episteme* 10(3), 241-268. ISSN: 1742-3600*

2012 **Jönsson, M. L.** & Hampton, J. A. (2012). The Modifier Effect in within-category induction: Default Inheritance in Complex Phrases. *Language and Cognitive Processes*, 27(1), 90-116. ISSN: 0169-0965

2011 | Hampton, J., Passanisi, A. & **Jönsson, M. L.** (2011). The modifier effect and property mutability. *Journal of Memory and Language*, 64, 233-248. ISSN: 0749-596X

Olsson, E. J. & **Jönsson, M. L.** (2011). Kinds of Learning and the Likelihood of Future True Beliefs: Reply to Jäger on Reliabilism and the Value Problem. *Theoria*, 77(3), 214-222. ISSN: 0040-5825

4. Research Review Articles

2011 | **Jönsson, M. L.** (2011) Direct Compositionality (by Chris Barker and Pauline Jacobson, Oxford University Press) *Language*, 87(1) 178-181. ISSN: 00978507

5. Books and Book Chapters

2012 | Hampton, J. A. & **Jönsson, M. L.** (2012). Typicality and Compositionality: The logic of combining vague concepts. *The Oxford Handbook of Compositionality* (eds. M. Werning, W.Hintzen, and E.Machery) 385-402. Oxford University Press. ISBN: 978-0-19-954107-2

8. Popular Science Articles

2010 | **Jönsson, M. L.** (2010). Thinking Outside the Laws of Thought. Published on the Atomium Culture Website: http://atomiumculture.eu/node/327 and through Atomium Culture distributed to sixteen major European newspapers (e.g. *Der Standard*, *El País*, *Postimees*.)

CV

CV - Martin Jönsson

Namn: Martin Jönsson
Födelsedatum: 19770614
Kön: Man
Land: Sverige

Dr-examen: 2008-10-04
Akademisk titel: Docent
Arbetsgivare: Lunds universitet

Utbildning

Forskarutbildning

Examen	Organisation
Doktorsexamen, 60301. Filosofi, 2008-10-04	Lunds universitet, Filosofiska institutionen 500011

Utbildning på grund- och avancerad nivå

År	Examen
2005	60301. Filosofi, Magisterexamen, Lunds universitet

Arbetsliv

Anställningar

Period	Anställning	Del av forskning i anställningen (%)	Arbetsgivare
juni 2013 - Nuvarande	Lektor, Tillsvidareanställning	25	Lunds universitet, Filosofiska institutionen 500011
januari 2009 - maj 2011	Forskare, Tidsbegränsad anställning	100	Lunds universitet, Filosofiska institutionen 500011
september 2002 - augusti 2008	Doktorand, Tidsbegränsad anställning	100	Lunds universitet, Filosofiska institutionen 500011

Postdoktorvistelser

Period	Organisation	Ämne
juni 2011 - maj 2013	City University London	60301. Filosofi

Forskarutbyten

Period	Typ	Organisation	Ämne
juni 2009 - maj 2010	Gästforskare	Rutgers University	60301. Filosofi

Uppehåll i forskningen

Period	Beskrivning
2014-06-13 - 2014-08-31	Parental Leave
2009-08-24 - 2010-02-21	Parental Leave

Meriter och utmärkelser

Docentur

År	Ämne	Organisation
2013	60301. Filosofi	Lunds universitet, Filosofiska institutionen 500011

Bidrag erhållna i konkurrens

Period	Finansiär	Projektledare	Din roll	Delbelopp (kr)	Totalt belopp (kr)
2011 - 2013	VR - Vetenskapsrådet	Martin Jönsson	Projektledare	0	640000
2009 - 2011	VR - Vetenskapsrådet	Erik Olsson	Medverkande	0	1875000

Priser och utmärkelser

År	Namn på priset/utmärkelsen	Utfärdare
2008	King Oscar the IIs Prize for best Ph.D. Thesis in the humanities at Lund University 2008	Lunds Universitet

Publikationer

Publikationer - Martin Jönsson

Namn: Martin Jönsson
Födelsedatum: 19770614
Kön: Man
Land: Sverige
Dr-examen: 2008-10-04
Akademisk titel: Docent
Arbetsgivare: Lunds universitet

Publikationer är avstängt för Jönsson, Martin på den här ansökan.

Registrera

Villkor

Ansökan ska förutom av den sökande även signeras av behörig företrädare för medelsförvaltaren. Företrädaren är vanligtvis prefekten vid den institution där forskningen ska bedrivas men beror på medelsförvaltares organisationsstruktur.

Signering av den *sökande* innebär en bekräftelse av att:

- uppgifterna i ansökan är korrekta och följer Vetenskapsrådets instruktioner
- eventuella bisysslor och kommersiella bindningar har redovisats för medelsförvaltaren och att det där inte framkommit något som strider mot god forskningssed
- nödvändiga tillstånd och godkännanden finns vid projektstart, exempelvis avseende etikprövning.

Signering av *medelsförvaltaren* innebär en bekräftelse av att:

- den beskrivna forskningen, anställningen och utrustningen kan beredas plats under den tid och i den omfattning som anges i ansökan
- medelsförvaltaren godkänner kostnadsberäkningen i ansökan
- den forskning som utförs inom projektet bedrivs i enlighet med svensk lagstiftning

Ovanstående punkter ska ha diskuterats mellan parterna innan företrädaren för medelsförvaltaren godkänner och signerar ansökan.