Received: 19 December 2022 | Revised: 14 June 2023 | Accepted: 2 August 2023

DOI: 10.1111/josp.12545

SPECIAL ISSUE ARTICLE

JOURNAL OF SOCIAL PHILOSOPHY

Critical theory, ideal theory, and conceptual engineering

Andrea Sangiovanni

Department of Philosophy, King's College London, London, UK

Correspondence
Andrea Sangiovanni, Department of Philosophy, King's College London, Strand WC2R 2LS, London, UK.
Email: andrea.sangiovanni@kcl.ac.uk

I shall take critical theory to have four main characteristics. First, it is a form of social, cultural, or political critique whose aim is emancipation, where emancipation is liberation from oppression or injustice.[1] Second, as a result of its emancipatory focus, critical theory is *embedded*.[2] It is embedded in the sense that it is focused on understanding and overcoming forms of unfreedom that impinge on the lives of people now and around here. Third, as a result of being embedded, its critical component emerges from a diagnosis: how and why do currently existing practices oppress those enmeshed in them? Fourth, it is self-reflexive about the role that theory (including its own theorizing) might play in promoting and/or preventing emancipation. Critical theory is, in short, *emancipatory*, *embedded*, *diagnostic*, and *self-reflexive*.

Critical theory thus always begins from the particular lived experience of people here and now, and focuses on the social structures and practices within which particular choices are channeled and enabled. A critical theory might begin with puzzles such as the following: why do people's educational and professional choices still cluster along gender lines in ways that seem, at least at first glance, to disadvantage women?[3] Or: What might explain why many working-class voters resist egalitarian policy-making proposals that would seem, at first glance, to work in their interest?[4] Alternatively, critical theory might begin with values that have become particularly salient in contemporary social and political debates, such as solidarity, and ask things like: Why has solidarity become important here and now? What needs does its invocation respond to? What scope, if any, does its invocation have for improving things?[5] What might talk of solidarity hide or obscure? For the critical theorist, principles and theories are always enmeshed in particular constellations of power and counter-power; one must always be self-conscious about the needs and functions that principles, forms of consciousness, and theories are meant to meet (in the case of needs) and play (in the case of functions) in any given society. There is therefore no such thing as a division of labor between empirical and normative theorizing; there is no suggestion that the theorist should merely aim to justify political

principles and plans, and then leave empirical social scientists or policy analysts to implement them.[6] The empirical, descriptive, and diagnostic understanding of how the social world is, and how it constrains, provides the background against which it makes sense to ask and answer normative questions.

I will not attempt to specify critical theory any further. Under its heading, we can include classical Frankfurt-style Critical Theory, ideology critique, critical race theory, disability studies, post-colonial theory, and radical feminism, among others. But I do want to say something about the contrast between critical theory and ideal theory. It strikes me as misleading to say that critical theory and ideal theory are incompatible. There are important differences, but they are, I want to argue, matters of *methodology* and *emphasis* rather than *essential structure*. Indeed, I will argue that, to the extent that there are differences between the two, ideal theories should become more like critical theories. Indeed, *all* social and political philosophy should become critical theory. To motivate this argument, I will use some recent tools from conceptual engineering to suggest why the critical theorists' emphasis on the *functions* that principles, theories, and forms of consciousness play in the here and now along with close attention to the *needs* that principles, theories, and forms of consciousness are intended to meet are key to justifying *all* social and political principles designed to guide reform and transformation of current institutions.

1 | I

Ideal theories are theories that seek normative principles to guide the design of practices and institutions in idealized conditions. The conditions are idealized in the sense that they require strict compliance, the suspension of any concerns with feasibility, and a final end-state in which further reform or radical revision of principles is no longer necessary.[7] Already in this thin description, it might seem obvious that critical theory, with its emphasis on the here and now, and ideal theory, with its emphasis on idealization, must be fundamentally at loggerheads. But this is not true. Critical theories can be (and often have been) *utopian*. Good examples are Donna Haraway's classic and enigmatic "Cyborg Manifesto" and the feminist science fiction—including Ursula K. Le Guin's *Left Hand of Darkness* (1969)—she uses to imagine new and radical possibilities of gender and human-machine hybridism.[8]

But it would be equally misleading to call critical-utopian theories ideal (in the Rawlsian and post-Rawlsian sense). The radical possibilities for cyborg identities that Haraway imagines are not meant to be read as programmatic; they do not suggest concrete ways that, say, women (or men) should be (or become) here and now, or describe principles that are intended to form a regulative ideal for current social and legal institutions. The same goes Le Guin's Hainish world of the future, populated as it is by ambisexual Gethenians. Rather, both are meant to loosen the hold that current sex-gender ideology has on us by forcing us to reimagine what is possible. Reading Le Guin and Haraway makes us self-conscious; we might encounter resistance and difficulty in reconstructing what the Gethenians are like, or what a cyborg might be like, without relying on our own assumptions of how sex and gender work. By describing alternative worlds and realities, they force us to abandon standard dualisms (man/woman; natural/artificial; organic/inorganic), and to imagine radically plural and shifting identities.

But critical theories are not adequately characterized as *nonideal* theories either. This is for the following reason. Nonideal theories *presuppose* a prior account of ideal theory: nonideal theories are theories that specify what is required of us here and now *given* some account of

principles that would govern an idealized society. They mediate, as we might say, between the real and the ideal. But critical theories do not typically have a two-stage structure in which principles for idealized conditions are first justified, and only then applied to the world as it is. Critical theory rather works the other way around: while critique presupposes commitment to evaluative and normative standards (which make sense of the critique of current practices and institutions *as* critique), those standards are not worked out first as regulative principles for idealized conditions. They *inform* critique rather than *prescribe* particular policies or plans of action. This is particularly clear in Le Guin and Haraway (but also applies to other critical theory). In Le Guin and Haraway there is, for example, no attempt to describe how we might move from principles for idealized conditions to principles for the here and now; similarly, there is no program of institutional reform, no concrete institutional designs, and no theory of the second-best. But, because Haraway and Le Guin have the current sex-gender system (and accompanying dualisms) firmly in their cross-hairs, and because both aim to disrupt and subvert the idea that the sex-gender system is really as natural or inevitable as many believe it to be, they both still count as critical theory.

This is also the case for more standard critical theories. Adorno and Horkheimer's *Dialectic of Enlightenment*, for example, does not begin with principles for an ideal society. Rather, it begins with a critique of the dominant ideologies and practices shaping current capitalist societies (in particular, the dominant modes of instrumental rationality). But this is not to say that the critique contains no evaluative element. It draws on values such as spontaneity, creativity, and authenticity to inform and shape its critique of, for example, the culture industry.[9] The possibility of a different society is implied by the critique, but it is not spelled out in any detail. For Horkheimer and Adorno, this is by design: there is no way to know what such a society might look like until we create the preconditions for its existence. The glimpses we have of a society that transcends our own will always be imperfect because we have been profoundly shaped by the very society that is the object of the critique. Because there is no attempt to work out what the implications of principles designed to regulate ideal societies are for our current institutions or practices, it would be misleading to call theories of this kind "nonideal" in the Rawlsian sense.

The conclusion I want to draw at this point is that critical theory and ideal–nonideal theory are orthogonal to one another. They are not incompatible; rather, they simply proceed in different ways with different emphases. In the rest of the paper, I will go further: I will use some of the resources provided by recent work in conceptual engineering to argue that *all* social and political philosophy—including Rawlsian ideal theory—should be critical in the sense I have described, and where it is not, it is useful to see it *as if it were*. It should, that is, be *emancipatory*, *embedded, diagnostic*, and *self-reflexive*. From this point of view, the fact that ideal theories start with principles for idealized conditions is not a problem if the idealizations are reinterpreted as heuristic devices for a more context-embedded critique of currently existing societies, and if ideal theories are more explicit about the emancipatory role that principles and theories are expected to play within current societies.

2 | II

A typical picture of disagreement in standard social and political philosophy distinguishes between disagreement over what a concept—say, JUSTICE, FREEDOM, EQUALITY, HUMAN RIGHTS[10]—means, and disagreement about what each of these might require of us, or what makes each of

these things good. The former is taken to result, more often than not, in (insignificant) verbal dispute; the latter, more explicitly normative or evaluative task, is taken to be *substantive* and *interesting*. You say that FREEDOM is a negative concept, I say it is a positive one. You say that human rights are those moral rights we possess in virtue of our humanity; I say they are those morally justified legal rights that appear in a series of postwar international legal instruments. I say that the concept SOLIDARITY is used to pick out a particular kind of joint action; you say it is used to pick out forms of altruistic, pro-social behavior based on identification with others. It is unclear what is at stake in these kinds of dispute. The worry is that nothing is at stake. Critics just come to the table with slightly different working understandings of a concept, and so have in mind different topics (despite their use of the same term). On this view, the interesting questions are all normative or evaluative and *only come into view once we agree on one (or another) understanding of a concept*. We agree that we are focusing on FREEDOM as a negative concept, and ask: What kinds of (negative) freedom, if any, are valuable, or otherwise worth pursuing? Or we agree that we are focusing on the positive concept, and ask the same question. On a descriptive level, there seems to be nothing to adjudicate the dispute. Or, if there is, it seems to involve the wrong kind of inquiry. Social and political philosophy, it will be said, is not lexicography; we are not in the business of tracing everyday usages of terms and their associated concepts and meanings. This is a job best left to the editors of the OED.

This common picture is misleading for two reasons. First, if we focus on lexical concepts (namely mental devices of representation that are expressed by corresponding terms in a natural language), the distinction between *substantive* and (merely) *conceptual* disagreement is obscure. On the face of it, if we both have a shared view about all the empirical facts, and I say "this policy is just," and you say, "no, this policy is unjust," then we disagree about the extension and intension of "justice," and so also disagree about the content of the concept JUSTICE used to pick out instances. On the face of it, then, we *do* have a conceptual disagreement that is also substantive. If this is right, then in what sense do we share the concept JUSTICE, but disagree what makes something an instance of justice? How can the two come apart? More needs to be said. Note that it will just restate the problem to say: "What is being offered are two different *conceptions*, understood as interpretations, of a *single* concept." The conceptions in question pick out *different* properties; it then looks obscure, without further argument, how they can be different "interpretations" of a *single* concept. Why is this not an instance of a merely verbal disagreement? Given the different content of each proposed conception, what ensures that the two interlocutors are discussing the same topic?[11]

Second, and closely related, we can question the assumption that a concept must be shared in order for a disagreement to be genuine. This can be the case for at least two reasons (both of which are drawn from the conceptual engineering toolkit): (a) two people operating with different concepts might be *metalinguistically negotiating* which concept is better suited to a particular context[12]; (b) two people operating with different concepts might best be understood as attempting to capture an underlying kind (in which case the concepts can be meaningfully compared by assessing which one is better suited to unite paradigmatic instances into a type).[13] Exploring each possibility carries a hidden implication, I will argue, for why social and political philosophy ought to be *embedded* and *self-reflexive* in the senses discussed above (I return to *diagnostic* and *emancipatory* functions below).

Regarding (a). The mistake made in the common picture is to think that two interlocutors must *share* a concept before their disagreement can be meaningful. The common picture overlooks the possibility of what Sundell and Plunkett call a *noncanonical* dispute.[14] A *canonical* dispute is a dispute in which two interlocutors meaningfully disagree, the first asserting a

proposition that the other denies. To illustrate: "The cat is on the mat"; "No, the cat is not on the mat" is a typical *canonical* dispute, in which one party asserts the very same proposition that the other denies. In a non-canonical dispute, by contrast, the two parties meaningfully disagree but assert *different* propositions. Good examples of *non*canonical disputes involve context-sensitive, gradable adjectives like "spicy," "cold," or "tall." A newly arrived visitor to a remote outpost in Antarctica meets a long-term resident of the outpost. Looking at a thermometer with a clearly visible temperature, the visitor says "it is cold"; the resident says "this is definitely *not* cold." "Cold" is temperature below some threshold. The two interlocutors use very different thresholds in making their assertions. They are therefore using *different* concepts of cold to represent the temperature at the station, and so pick out different properties. Are they speaking past one another? Plunkett and Sundell argue that they are not. They are, rather, negotiating over which concept is appropriate to use at the outpost. The resident, that is, is communicating information about what "cold" *ought to mean now and around here* given average temperatures for the season, and so on.

The social and political philosopher can make sense of the disagreements mentioned above using the same strategy. When two interlocutors disagree about whether freedom is negative or positive (or something else), they can be understood not as disagreeing about the everyday definition of a word, or as merely talking past one another, but as metalinguistically negotiating which concept of freedom ought to be used now and around here (just like "cold" in the Antarctic).

Understanding the disagreement in these terms has, I want to argue, an important upshot for social and political philosophy. Most treatments on the nature of freedom, human rights, solidarity, and so on, do not spend much time delineating what particular contexts they are proposing their concepts *for*. The most one gets in discussions of freedom, for example, is that one aims to provide an analysis of *political* freedom (as opposed to an account of free will). But there is rarely discussion of what is meant by "politics." What specific political circumstances does one need a concept of freedom for? For what uses? Is this a concept of freedom for the carceral system? For establishing a constitutional list of rights? For evaluating which regimes secure greater political freedom overall than other regimes (and what is the *point* of such an evaluation in the first place)? This background is taken for granted. Yet, if we take the idea of metalinguistic negotiation seriously, specifying the context is essential. This is because it will determine the criteria to adjudicate which concept is most useful, or appropriate, for that context. To draw the analogy, it matters for determining what concept of cold we should use whether we are talking about Antarctica or Los Angeles, and it matters, too, what we are meant to be doing (are we leaving for an expedition to rescue someone? Are we planning a party?).

This contextualization implies that, on the metalinguistic model, social and political philosophy should be embedded and self-reflexive in the senses described in the introduction. Theorizing must begin with current practices, and must make sense in terms of those practices. We need to know what contexts our substantive conceptions of justice, democracy, and so on, are meant to guide before we know which one we should choose. Just as it would be absurd to claim that there is a single, all-purpose concept of cold that we ought to use in the Antarctic and in LA, it would be just as absurd to claim that there is a single, all-purpose concept of justice, democracy, and so on, to use in every context in which it makes sense to apply the corresponding term. Different proposals regarding the nature of justice, democracy, equality, and so on, are intended, on this picture, as contributions to what we believe justice, democracy, solidarity, and so on, *ought to be for us now and around here*. Just as in the Antarctica example, what will determine the best solution will depend on contingent features of the practices and

institutions to which our conceptual proposals are meant to apply, and on the role we think our theories should play with respect to those practices and institutions. Once again, this is not necessarily nonideal theory. The concepts proposed might be intended, like Haraway or Le Guin, as challenging us to imagine different possibilities, and so put into question the seeming inevitability of our current sex-gender system, rather than guide, say, decisionmakers within some particular institution. But, for all that, the concepts still only make sense as useful or appropriate or fitting against this background.

Note that, on this picture, proposals regarding what justice, democracy, equality, and so on, ought to mean now and around here are not assumed to track mind-independent and practical-interest-independent properties. The best proposal will be best, not because it tracks what justice, democracy, and so on, *really* is, but because it is the most appropriate or fitting or adequate given the particular purposes we want it to play here and now. This does not, however, commit one to moral anti-realism. It might be that, in evaluating which concept of justice, democracy, freedom, and so on, is best to use now and around here, we should appeal to higher-level, more basic values and principles that are fully mind- and practical-interest-independent.[15] We might ask: Which concept best promotes the good? Which one, when adopted, would help to sustain relations that no one could reasonably reject? Which one best protects our inviolability as morally autonomous agents? Disagreement about the basic standards of evaluation might themselves be subject to metalinguistic negotiation, or they may not (I return to this below). On this view, we might think of the lower-level, less basic conceptions of democracy, equality, freedom, and so on, as constructed from such higher-level principles and values conjoined with an account of the role the lower-level conceptions are meant to play here and now. And to know the role they *should* play, we also need to know something about the role that such concepts *do* play. To put it in the terms with which we began: there is no division of labor between normative political theory and social science.

We can go further. The contextualization needed for metalinguistic negotiation leads naturally to social and political philosophy that is not only embedded and self-reflexive, but also emancipatory and diagnostic. Insofar as social and political philosophy is embedded and self-reflexive, it arises as a *response* to forms of impingement in the here and now. On this view, theories of justice, democracy, and so on, are needed *only* because of the oppression that injustice, authoritarianism, and so on, bring in their train. They matter because of their transformative potential. And, if they are intended to be transformative, it is also natural to ask what obstacles they face in being realized (and what the dangers might prevent their realization). What might an emphasis on democracy obscure? How can a utopian critique inform?

A good example is the current dispute about what the term "woman" (and its associated concept) should refer to now and around here. Take an advocate of the idea that WOMAN ought to pick out any person that (roughly) self-identifies as a woman, and compare that to a proposal that WOMAN should more clearly and unambiguously be used just to refer to biological sex (and so be synonymous with FEMALE).[16] And suppose that both disputants agree that neither proposal is an analysis of the concept that most people in our society currently use to pick out women. Here it seems obvious that the two disputants do *not* share a concept of woman. This kind of dispute could be fruitfully interpreted, I am suggesting, as a metalinguistic dispute. And it also seems clear that the dispute could still be meaningful, i.e., not an instance of a merely verbal dispute. What would make sense of the dispute would be a deeper analysis of what roles each of the disputants believe that WOMAN should play in our society (including its role in the law and in society more generally). Of course, the ethical and critical import of the dispute becomes evident when we think of feminist struggles to liberate women from subordinating roles and in

light of trans-activism: What concept of woman would best promote feminist ends? What concept might best promote solidarity? Should the concept of woman be extended to include trans-women? How? How do current classification practices create unfreedom? How might new usages (if at all) liberate? On the metalinguistic proposal, we could think of disputes on the nature of justice, democracy, equality, and so on, *in the same way*: different proposals to be evaluated in light of their envisioned role in particular contemporary social, political, and legal struggles. They must be, therefore, at once, emancipatory, embedded, self-reflexive, and diagnostic.

Note that the metalinguistic negotiator can grant that a term like "freedom" (and the concept it expresses) might have a more abstract meaning (what Kaplan calls a *character*—a function from a context to a determinate proposition[17]), while also holding that the concept associated with it will only have a determinate content in a particular context. Gradable adjectives work in a similar way. There is a general, all-purpose meaning of "cold" that is not relative to context, namely *temperature below some threshold*. But expressions that use the term will be incomplete—they will fail to express a determinate proposition—until information about the context fixes the threshold. Only then will expressions that use the term be truth-evaluable. And so it might be, on the picture I am proposing, with terms like "freedom" (and its associated concepts). "Freedom" may have a general, all-purpose character—we may use a modification of MacCallum's well-known triadic definition—"absence of restraint from *x* for agent *y* to do *z*"—but expressions in which "freedom" figures will not be truth- or appropriateness-evaluable (for a given context and purpose) until the variables in the definition receive a value.[18] This is true even for more complex expressions such as the following: (in a discussion of the carceral system) "it would be better if prisoners had more freedom." To evaluate the truth of the statement, we need to know, among other things, (a) what *kind* of freedom is at stake, (b) what the *point* of making assertions of this kind is in this context for this purpose (are we, e.g., legislators seeking legal reform, philosophers assessing the justice of a carceral system, etc.)?, and (c) what other values, principles, standards and so on, are in play (why is this kind of freedom worth talking about in this context and for this purpose)? On this picture, there is no such thing as *the* concept of political freedom. There are many possible concepts of political freedom, each suitable for a different context and a different purpose (and some, of course, unsuitable for any context).[19]

So far, in explicating metalinguistic negotiation, I have used the example of context-sensitive expressions, and suggested that concepts like FREEDOM could also be understood as having a more abstract character that picks out particular objects only in the presence of a determinate context. But metalinguistic negotiation can also be appropriate for concepts that are *not* context-sensitive. An example due to Peter Ludlow, used also by Sundell and Plunkett, is ATHLETE in the context of a discussion over whether a particular winning, graceful, powerful, and hard-working race horse ("Secretariat") is an athlete.[20] One interlocutor might deny what the other affirms about the horse, even where they agree on all the empirical features. And this dispute might well be genuine (depending on what rides on the affirmation). In this case, it would be implausible to argue that "athlete" is a context-sensitive term with a character that gets filled in depending on the context; rather, what is at stake are two distinct concepts that do not share a character (though they share an overlap in many paradigmatic instances). This matters for our purposes because it allows us to retain a metalinguistic analysis even where it seems implausible to argue that some concept within social and political philosophy (such as JUSTICE) is context-sensitive in the way that COLD is.

One might worry that taking this metalinguistic route leads to fragmentation. To make this point, we can shift examples. What if (as I have argued elsewhere[21]) we take the concept HUMAN RIGHTS to have a determinate content only once a particular political context has been specified? Suppose, as I have been suggesting, that we take the resolution of a disagreement between someone who asserts a Legal Conception of human rights[22]—such that human rights are those morally justifiable individual legal rights whose systematic violation would warrant some form of international legal remedy—and an Orthodox Conception[23]—such that human rights are moral rights we possess merely in virtue of our humanity—to depend on the context in which the respective concepts of human rights are offered. On this view, each might be appropriate for a different context: an Orthodox Conception might be more appropriate for the work of NGOs, and a Legal Conception more appropriate for foreign policymakers. Therefore, we might have a human right, say, not to be lied to according to the Orthodox Conception, but not according to the Legal. This leads us to the odd-sounding conclusion—if each concept is appropriate for a different context—that we both have a human right not to be lied to and do not have it. But aren't human rights meant to be universal? And how would two interlocutors across contexts be talking about the same thing?

The answer to both questions is *internal* to the dispute. The worry about fragmentation is motivated by the thought that there *ought to be* a single concept of human rights. To allow for diversity would make the concept incapable of serving the main functions that human rights are there to promote in the first place. But this is to appeal to just the kinds of considerations just mentioned for evaluating concepts: What purposes ought the invocation of human rights serve in the main practices in which the term has found a place (answer: one of the main purposes is to promote the universal scope, political power, and reach of human rights)? Whether pluralism in this context is, or is not, appropriate depends on the particular features of the case we are considering.

Before turning to (b), it is useful, by way of contrast, to see the effect of applying the metalinguistic strategy to Rawlsian ideal/nonideal theory. Suppose we are contrasting Rawlsian JUSTICE with Nozickean JUSTICE and take a metalinguistic reading. Our two interlocutors disagree fundamentally about paradigmatic instances of justice and injustice ("considered convictions"). We do not, on this reading, assess which one captures what justice *really* is. Rather, we first ask: What purposes is each proposal meant to serve? How is it meant to illuminate or guide our current practices? What social phenomena is each proposal intended as a response to? The social and political circumstances in which each proposal is meant to apply take central stage (recall Antarctica). It will be misleading, from this perspective, to say something like "Rawls's Two Principles are intended to govern a well-ordered society, and so it is a further question (of nonideal theory) to determine how and whether they apply to our society here and now." That would be like abandoning the only field in which Rawls's proposal could be compared with Nozick's (unless one was truly interested in the entirely hypothetical question—"which set of principles are better suited to a well-ordered society?"). Rather, to assess them, including their assessment against higher-level moral principles, requires much further historical, social, cultural, and political contextualization. What kinds of contemporary social, cultural, political, and economic phenomena does each theory encourage us to focus our attention on?[24] What is the effect of viewing a society in terms of principles that are designed to govern a well-ordered society? Who, and in what contexts, is intended to adopt the concept of justice that each proposes? How does each theory aim to emancipate or free from oppression? Adopting the internal perspective of the theory, what forms of oppression does it encourage us to focus on (which ones does it obscure)? Moral considerations, of course, will matter in such assessments but only

against a much thicker account of our practical interests in theorizing about justice here and now. We ought to read, interpret, and assess them, that is, as much more like critical theories than one might at first take them to be, which is to say that we ought to read, interpret, and assess them as if they were *emancipatory*, *embedded*, *diagnostic*, and *self-reflexive*. In Rawlsian terms, the ideal component of the theories would be assessed by their nonideal functions, rather than the other way around.

3 | III

Regarding (b). So far I have discussed how metalinguistic negotiation can provide a powerful tool from the conceptual engineer's toolkit for how to think about even the most ideal forms of social and political philosophy *as* critical theories. The second resource offered by conceptual engineering has a more externalist character. Instead of thinking of the disputes mentioned above as metalinguistic negotiations over which concepts are most appropriate for different contexts, we can think of them as trying to capture essential features of a set of social and political practices—we can think of them, that is, as attempts to characterize the type that unites a set of paradigmatic instances of freedom, human rights, and so on, into a *social or political kind*.[25] The idea, roughly, is that the concept figures in and organizes reflection, judgment, and action on a set of social and political practices that it, at the same time, refers to. The practices and their associated beliefs, norms, patterns, values, dispositions, discourses, and so on, form part of what Geuss calls a "form of consciousness," or a widely shared, systematically interwoven conceptual scheme that shapes the background culture of a society.[26] There are practices of (declaring, advocating for, challenging) human rights, practices of (expressing, exercising, fighting for) freedom, practices of (demanding, organizing, acting in) solidarity. Over time, the practices coalesce into a unified kind that the associated terms refer to; the kind then takes on a life of its own that outstretches any one pattern of usage.[27] We can then study the kind and learn new things about it; the definitions of the terms we use to refer to those practices, once they are up and running, can fail to capture what unites the instances into the kind.

The central target concept we deploy in categorizing these practices is an attempt to pick out, on this reading, what Searle calls an "ontologically subjective, but epistemically objective" kind.[28] It is ontologically subjective because, as a social kind, its existence depends constitutively (rather than merely causally) on our having appropriate attitudes toward it (e.g., beliefs and intentions).[29] For example, money (*qua* legal tender) and marriages depend *constitutively* on our attitudes toward them; without those attitudes money and marriages would cease to exist as money and marriages. They are ontologically subjective. Certain breeds of dog, polyethylene, and plutonium, on the other hand, depend, for their existence, on our *causally* bringing them into existence. But once they have been brought into existence, they no longer depend on our attitudes to continue existing. Therefore, they are, like tigers, mountains, and water, ontologically *objective*. Social kinds like money and marriage are, despite being ontologically subjective, epistemically *objective*. We can ascertain, that is, whether they are true and false independently of people's attitudes or opinions. The truth of statements about what is, and is not, money (marriage, etc.) depend on a complicated set of satisfaction conditions that are independent of any one individual's beliefs about them.

The phenomenon of solidarity can be treated as a social kind in this sense. On this understanding, an account of the concept of solidarity is an attempt to capture the kind. In reconstructing what solidarity is, we might proceed in the following way. SOLIDARITY becomes,

in the late nineteenth century, a central concept for thinking about the social bonds that might tie together a society—including a willingness to stand by other members in difficulty—once traditional ties of kinship, church, and status have broken down. The concept emerges in a number of contexts, most prominent of which are workingmen's societies and professional groups within civil society (as in Socialism) or parties attempting to redefine the underlying social contract supporting the state or nation in an era of growing inequality (as in Solidarism and nationalism). It also, at roughly the same time, becomes prominent in Christian attempts to promote a meaningful form of human fellowship among the classes (i.e., workers with their employers, the rich with the poor, peasants and landowners) in distinction to both socialist "collectivism" and liberal "individualism." A *theory* of solidarity, in turn, aims to provide an account of the type that unifies the paradigmatic instances of solidarity into a social kind. It aims to provide, that is, an account of the real-world phenomenon our concept(s) of solidarity are trying to latch onto (as are the concepts used historically by each of the individuals that made the phenomenon salient).[30] As in the previous cases we have discussed, we may decide that there are *different* concepts of solidarity associated with different phenomena. Or one might try to argue (as I have elsewhere) that solidarity is a single phenomenon with an underlying structure that can be captured by philosophical-cum-historical-cum-sociological analysis; the unified character of the kind means that it can be used fruitfully in inductive, explanatory, and normative guises in sociology as well as social and political theory. As a social kind, we can study it, and learn more about it (including more than is contained within the concepts currently deployed to refer to it).[31]

What is the difference between this externalist picture and the one provided by the metalinguistic interpretation? The most important difference is that the externalist reading adds an epistemic dimension to evaluation lacking in the metalinguistic one. In a dispute between two interlocutors, each of which uses a different concept of solidarity, metalinguistic negotiation encourages us to adjudicate on the basis of the purposes we want the concept to play in a specific context. The reasons for preferring one concept to the other do not depend on which one is better able to cut the (social) world at the joints; the standards for evaluation are practical rather than epistemic. While one concept might be, all else equal, more coherent, clear, or determinate than another (thus giving one some epistemic reason to prefer it), one concept cannot be preferred to the other because it captures a phenomenon more *accurately* than the other. For example, in the "cold" example, the visitor says something *true* when they use "cold" to mean "temperature below 0 degrees Celsius" and say "it is cold," and so does the resident when they use "cold" to mean "temperature below −10 degrees Celsius" and say "no, it is not cold." The same thing holds when we are contrasting two different accounts of WOMAN. The dispute, as interpreted by the metalinguistic reading, does not turn on whether one or the other person says something false; it turns on which one is using the most useful concept given the context. When the resident denies that it is cold, they are attempting to communicate information about the standards that govern judgments of cold *in this place*. This is why Sundell and Plunkett refer to disagreements of this kind as *noncanonical*.

On the other hand, if one were to give an externalist reading of the dispute, then the visitor, when they utter "it is cold," would be saying something *false*; the resident, on the other hand, would be saying (we suppose) something *true*. The particular property associated with use of the term "cold" in that context is set by objective features of the context (not by the mental states of the interlocutors)—hence externalism. And so it is with our analysis of solidarity. In evaluating two different, competing theories of solidarity, one can ask: Which one is better at carving the (social) world at the joints?[32] So someone who uses a concept of solidarity that picks

out all and only instances of pro-social behavior based on identification, and so who believes that returning a lost wallet to someone is an act of solidarity, might be saying something false—false if, say, the history, values, purposes, and practices centrally associated with solidarity give the phenomenon a structure that excludes unilateral forms of altruism, however, motivated. The history, purposes, values, and practices that give solidarity its shape set the kind that our concept is trying to latch on to. But we may fail in latching onto them.

We can give a similar reading to the dispute about human rights. Rather than thinking of proponents of the Legal Conception and proponents of the Political Conception as both saying something true about human rights, and then selecting between them solely on the basis of their usefulness in different contexts, we can think of theories of human rights in a more externalist way. On this reading, we conceive of different theories of human rights as attempting to track the practice (or practices) of human rights. The theories attempt to pick out the core, structured, typical, and interwoven aspects of the practice that unify human rights into a social kind. Whether the theories are successful turns on objective and morally salient features of the practice, rather than the usefulness of a particular concept in particular conversations. So it will often be true, when there is a dispute, that someone using one concept in a given context is saying something false when they assert the existence of a human right whereas the other is saying something true when they deny it. The truth conditions of a statement like "the right to education is not a human right" are set, on this understanding, *not* by whatever concept of human rights the interlocutor has in their head, but by whatever human rights really are in that context (which can depend both on the purposes they serve *and* the moral values and principles they encode, and so on).

Haslanger's account of WOMAN fits this model.[33] According to Haslanger, a woman is, roughly, a person who is assumed both to have female bodily features and to be subordinated. Haslanger's proposal for how to define "woman," she argues, is an attempt at capturing an underlying social kind set by our current sex-gender practices. Competing proposals are to be assessed according to whether they succeed at capturing this kind, including the roles played by sex-gender categorization in our society. If her account is successful, it lays bare the fact (if it is a fact) that our categorization practices not only serve to pick women out on the basis of (perceived) female biological properties, but also serve to subordinate through the assignment of gender-based social roles and expectations. By calling attention to these social facts, her account is therefore intended to be emancipatory, since it is intended to get us to question whether women, as understood by the account, ought to exist (given the fact that subordination is required for women to exist).[34]

The same strategy can be deployed for more abstract concepts like JUSTICE. On an externalist reading, judgments of justice are treated as modes of public contestation whose point is to reform, sustain, and replace specific normative orders and social structures. Concepts of justice are attempts to track those practices; they must be evaluated, then, by whether and how they unify those practices into a morally salient type, which will include evaluating whether the concept, as articulated in a *theory* of justice, serves the purposes for which we need assessments of justice in the first place, can enter in the right kinds of inferential relations with appropriate reactive attitudes and responses, and is focused on structures, norms, and institutions that call out for reform. Proposals for what justice is must then be assessed, as in the metalinguistic interpretation, in terms of the target practices, associated norms, and the reactive attitudes involved. And so we come to the same conclusion, namely that, even within an externalist framework, social and political philosophy ought to read, assessed, and interpreted as emancipatory, embedded, diagnostic, and self-reflexive. The difference is that we do not assess

proposals solely for their usefulness or appropriateness (as in metalinguistic negotiation), but also in terms of whether they are better at unifying those practices, norms, and attitudes into a single morally salient type. On this reading, a concept of justice can fail to represent the underlying social and morally salient kind adequately; if it does, then the particular judgments that express the concept can be false in a way that the metalinguistic reading resists (where they can only be more or less useful or appropriate).

To sharpen the contrast between metalinguistic and externalist views, and to bring out their critical normative potential, it is useful to highlight the distinction between three different ways in which the kinds of disputes we have been considering involve moral and evaluative considerations. So far, I have referred to the concepts discussed—SOLIDARITY, JUSTICE, HUMAN RIGHTS, and WOMAN—as "morally salient," but I have not discussed in any detail how moral and evaluative considerations enter the picture on either view. Social and political concepts can be either (a) descriptive but normatively dependent,[35] (b) moralized and thick, or (c) moralized and thin. Descriptive but normatively dependent concepts, like WOMAN, are not *moralized*, by which I mean that the literal content of the concept is not *directly* fixed by any moral or evaluative facts (or attitudes). There are just and unjust, good and bad, morally rightful and morally wrongful women. WOMAN does not pick out any evaluative or moral facts (or attitudes) in the world; woman is, therefore, a *descriptive* category. Moral and evaluative facts (or attitudes), however, still matter *indirectly*. The category, woman, includes moral implications and normative expectations ('if one is a woman, one has a duty to..."), but these role obligations are delineated and fixed *conventionally*. One does not need to make an evaluation of whether some purported duty is *truly* a duty or not to determine whether someone is a woman. Someone is a woman, now and around here, as long as she is *expected* to comply with whatever duties and obligations are thought to apply to women as women. Woman, as a result, is a *normatively dependent*, but not a moralized, category. As I have alluded to above, I believe the best characterization of solidarity makes it out to be descriptive but normatively dependent in the same way. While people acting in solidarity, for example, have morally laden expectations of one another that are constitutive of the phenomenon, people can count as acting in solidarity *even if the morally laden expectations do not spell out genuine duties or moral reasons to do anything*; neo-Nazis, on this account, can act in solidarity.

On the metalinguistic reading, then, properly moral and evaluative considerations enter at one remove: they help to adjudicate which word-concept pairing one should choose (including whether to choose descriptive but normatively dependent concepts, or moralized versions of the same concept). Which word-concept pairing regarding woman (or solidarity), for example, will secure the morally and evaluatively best outcomes if generally adopted? What is the moral or evaluative point of invoking the term in a given context (and: what should it be?)?[36] On the externalist reading, properly moral and evaluative considerations also enter at one remove, but in a different way. Here the question is: Which concept best captures the practices in which the category of woman/solidarity plays a central role? Because (we are assuming) woman and solidarity are normatively dependent, moral and evaluative considerations will influence how the category is shaped by participants within those practices. What role do, say, moral expectations of reciprocity play in practices of solidarity? What particular duties and obligations are women as women supposed to fulfill? The externalist, that is, takes a more sociological, and less prescriptive, view of the moral and evaluative considerations in play.

But how does the distinction play out when *moralized* concepts are at stake? Let us take thick moralized concepts first. Thick moralized concepts are thick because they combine descriptive and moral/evaluative components.[37] I have characterized justice as a thick

moralized concept. A theory of justice, on this view, is not just a view about what is right and wrong *simpliciter*. It is a theory of right and wrong *for* a given set of practices *in* a given social, political, cultural, economic context *against* a background of appropriate reactive attitudes and actions. On a metalinguistic reading, what is being negotiated between two parties to a dispute about "justice"/JUSTICE is not only the content of particular principles and reasons that we ought to recognize, but also what role or function "justice"/JUSTICE *should* play now and around here, what *should* follow from a judgment of "injustice"/INJUSTICE (and what relation such judgments have to the particular *form* of the society in question), and what role it *has* played. Moral and evaluative considerations come in at one remove, and serve as standards for evaluating and criticizing the various roles and contents envisioned for "justice" now and around here.

On an externalist reading, the emphasis is on the current practices and contexts in which people now and around here make judgments of justice. In the first instance, the question is not: What contexts/institutions/practices should "justice"/JUSTICE inform? Rather the question is: How *does* justice—as a social and moral phenomenon (hence disquotationally)—enter into and become relevant in current contexts/institutions/practices? What functions does it play in our society? Only once this content is fixed does the externalist turn to the question: What does justice morally require of us *in* and *for* those contexts/institutions/practices (whatever they are)? In the former case, external *social* factors determine what the function and shape of judgments of justice is. In the latter case, external *moral* factors determine what the content of justice is (given its functions in this society).[38] This marks a key difference from the metalinguistic view: whereas the metalinguistic reading treats justice as a set of proposed rules of social regulation which are negotiated in terms of other moral, practical, theoretical values, the externalist reading treats different theories of justice as attempts to latch onto what justice really is now and around here. For the metalinguistic reading, the point is not to determine which conception of justice is *true* (recall that every proposal is treated as generating true propositions about justice) but which conception is more *useful*. By contrast, for the externalist, the point *is* to determine which among competing conceptions has the more accurate picture of what functions justice plays, and what content it is has given those functions.

The fact that moral values play a role in fixing the content of justice as a kind may make it seem as if the externalist must be committed to moral realism. Justice as a kind would be composed of both social *and* moral facts. But this is not required. The externalist (in the same way as the metalinguistic negotiator) can leave open what the best semantics and metaphysics is for the moral component of justice. If one is, say, an expressivist like Gibbard or Blackburn, one can simply say that the moral component of justice is fixed by whatever the best system of attitudes, plans, or other conative attitudes governing what to do is (given the social function of justice), and that we should take a deflationary stance regarding claims about the truth of moral statements.[39]

This takes us to the last category, namely that of *thin* moralized concepts. These are concepts that have *no* descriptive component. They are exclusively normative. For these concepts, metalinguistic negotiation breaks down. Metalinguistic negotiation breaks down because, at least once one is dealing in the most basic, higher-level normative terms, there is no other perspective from which to evaluate them for their usefulness or appropriateness. It makes sense to negotiate different concepts of justice by asking "which one ought we adopt (from some other normative standpoint, independent of justice)?" But what about "ought" itself? Can we negotiate which concept of "ought" to use in some context? But then we would need to ask—'Which concept of "ought" ought we to use?"—in which case we would be moving in a circle.[40] Externalism for the most basic terms, similarly, provides no foundation for the claim that theorizing

the most basic moral terms should be emancipatory, embedded, diagnostic, and self-reflexive. I want to concede this point. It is enough for our argument if we have shown that, on either the metalinguistic or the externalist picture, theories of thick moralized concepts and descriptive but normatively dependent concepts—which are the most important concepts in social and political philosophy—must be emancipatory, embedded, self-reflexive, and diagnostic. This is as it should be: the most basic moral terms are needed precisely to provide a higher-level perspective from which to survey our myriad practices and institutions (including the concepts operative in them). It is also worth noting, however, that there will be very little to say about the content of such very basic normative terms *without* specifying some context of assessment. On the picture I have been developing, we deploy them, not as core structures, but as scaffolding on which to do theory that is emancipatory, embedded, diagnostic, and self-reflexive.

I end this section by reminding the reader that my aim is not to adjudicate between the metalinguistic and the externalist readings. It may be, as we have seen, that they are each adequate for different kinds of disputes. It also may be that, in some cases, it is appropriate to metalinguistically negotiate whether an externalist interpretation is itself the most useful for some given end.[41] (Indeed, this may be a way of reading Haslanger's more recent claim that different concepts of woman are appropriate in different contexts—in some contexts, her externalist version might be more adequate for social critique; in others, a concept that allows more scope for self-identification may be (as in debates about trans-persons).)[42] It is enough here if we see how these different tools provided by the conceptual engineer can provide a richer understanding of central disputes in social and political philosophy, and, more importantly, how they suggest ways in which social and political theories in general ought to be read, interpreted, and assessed as if they were all critical theories.

4 | CONCLUSION

One of the main distinguishing characteristics of critical theory (as opposed to mainstream political and social philosophy) is that it begins its theorizing *not* with an attempt to work out the content of abstract moral principles or values in idealized conditions, which are then applied to different contexts, but with our social practices now and around here. This starting point will include the moral, political, and social values and principles that form part of day-to-day life, and that we use to organize and coordinate our life together. And it will also include diagnosis of areas of social and political life—structures, norms, institutions, patterns of action—that undermine our ability to live freely (broadly understood). How do current social structures, institutions, and self-understandings make us unfree? How does the current form of consciousness within a society disguise oppression within those practices? What might it be like to rework or reject major supporting planks of the reigning form of consciousness? How can self-conscious and reflexive theorizing contribute to the process of emancipation? This is opposed to social and political philosophy that takes itself to characterize what, say, justice or freedom or solidarity is *independently* of any contingent social practices or particular historical contexts or local practical aims—what justice, freedom, solidarity, are, as it were, *period*—or, alternatively, that takes itself to characterize what justice, freedom, and so on would be if we could remake the social world from scratch. My main aim in this paper has *not* been to argue that such theories must therefore be mistaken. It is not, for example, as if principles defended in this way are (necessarily) false. Rather, my claim was that theories that proceed in this way are incomplete (and in some cases misleading) because they leave out a crucial dimension of

evaluation necessary for measuring the success of any theory and for adjudicating between them (given what I have said about what makes disagreement regarding political concepts meaningful).

To make this point, I began by questioning a common picture of disagreement in social and political philosophy. The common picture draws a distinction between *conceptual* and *substantive* disagreement, and assumes that conceptual disagreement is merely verbal: interlocutors must share a concept before they can disagree substantively about what it requires. I have argued that this need not be the case: disagreement in concepts can be meaningful if we assume either that interlocutors are metalinguistically negotiating which of many possible concepts of, say, freedom, democracy, solidarity, and so on, is most useful or appropriate in a particular context, *or* that the concepts used by interlocutors are attempts to track an underlying morally salient social kind. Exploring each possibility had a surprising upshot: to avoid pointless verbal disputes and empty speculation, there is a sense in which all social and political philosophy must become critical theory. Identifying the content, grounds, and scope of even basic social and political values and principles should not proceed entirely a priori. To resolve disputes between different proposals for how to understand a given concept requires us to look to the practical context in which the values and principles are meant to do their work. We need to ask: what is the *point* of the concepts representing the values and principles in our discourse and practice? What work are they meant to do? Why is it pressing to identify and coordinate around a determinate content? How did the concepts become dominant in social and political discourse, and what ends do they serve? What further values might be relevant in evaluating them? What social practices are these concepts meant to track? Only once we have such further ends, values, and practices in view can we specify the content of the concepts we are engaging, and hence only then can we make a substantive argument regarding what justice, solidarity, freedom, and so on, is. If we do not, then, once again, we are either running in place, or hiding (whether consciously or unconsciously) an agenda.

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest.

ORCID

Andrea Sangiovanni https://orcid.org/0000-0002-1915-3969

ENDNOTES

[1] Cf., Horkheimer, 1972, p. 244.

[2] Some want to say that it must always latch onto current *social movements*. See, for example, Haslanger (2012). I do not think this is necessary: a critical theory could be emancipating even if it does not lead to the development of a social movement. It might, for example, simply get one to rethink central categories that one uses.

[3] Cudd, 2006.

[4] (Horkheimer, 1988, pp. 188–252).

[5] This historically and socially situated focus of critical theory in this broad sense is central to contemporary theorists as diverse as Geuss (1981), Benhabib (1986), Shelby (2009), Jaeggi (2014), Forst (2017), Young (1990), Coulthard (2014), and Williams (2002).

[6] Cf., Swift, 2008.

[7] Valentini, 2012.

[8] Haraway, 1991; Le Guin, 1987.

[9] Ch. 4, (horkheimer dialectic).

[10] As is standard, I will use small caps to indicate the concept—for example, FREEDOM, indicating the mental device of representation that we use to sort information and categorize the world— quotation marks to refer to the word—for example, "freedom"—and no punctuation or caps to refer to the object of representation or reference—for example, freedom.

[11] Perhaps sufficient overlap in paradigmatic instances? (But there may not be much overlap, and overlap is not sufficient either—consider the overlap in paradigmatic instances between the concept of Berliner and the concept of German.)

[12] See, e.g., Plunkett & Sundell, 2013a.

[13] See, e.g., Haslanger, 2012; Schroeter & Schroeter, 2014.

[14] Plunkett & Sundell, 2013b, p. 260.

[15] The ultimate moral principles and values might, alternatively, receive an anti-realist treatment. The point is that the view under discussion here does not commit us metaethically to any particular theory.

[16] Cf., e.g., Hall, 2009; Jenkins, 2016; Lawford-Smith, 2022.

[17] Kaplan, 1989.

[18] MacCallum, 1967.

[19] Cf., the picture of different concepts of morality that Nietzsche contrasts in the *Genealogy of Morality*. See also Geuss, 1999, p. 168ff.

[20] Ludlow, 2008; Plunkett & Sundell, 2013a, p. 16ff.

[21] Sangiovanni, 2017, Ch. 4.

[22] See, e.g., Buchanan, 2013.

[23] See, e.g., Griffin, 2008.

[24] Cf., Forrester, 2019.

[25] On social kinds, see Searle (1995), Hacking (1999), Haslanger (2012), Khalidi (2015), Mason (2016), Godman (2020).

[26] Geuss, 1981, p. 10.

[27] See, e.g., Hacking, 1999 on interactive kinds.

[28] See, most recently, Searle, 2010, pp. 17–18.

[29] I draw the distinction between causal and constitutive dependence from Boyd (1989).

[30] Cf., the externalist account in Schroeter and Schroeter (2014).

[31] For an analogy, see, for example, Griffiths discussion of emotion as a normative and social kind (though not a natural kind) in Griffiths (2004).

[32] This does not require that the "joints" be precise. There can be borderline cases. The important thing is that the core features of the phenomenon be woven together in structured, typical, and recurring patterns across paradigmatic instances. See Boyd (1991) on natural kinds as clustered properties.

[33] See, more recently, Haslanger, 2020 in which she places her Haslanger, 2000 in the wider context of feminism and also current debates on trans issues.

[34] The aim of feminism, Haslanger writes, is to make it the case that "there will no longer be men and women" (Haslanger, 2012, p. 366).

[35] I am indebted to discussion in (Forst 2017), though the use I make of the term is different. I note here that although there are descriptive social concepts that are *not* normatively dependent (e.g., recession, inflation, voting, neighborhood, etc.), these will not be the focus of my attention here. (Each of those are not normatively dependent because people's taking themselves to have moral or evaluative reasons to do *x* are not part of their content.)

[36] For a good account of how normative and evaluative considerations enter into metalinguistic negotiation more generally, see Thomasson (2017), Thomasson (2020).

[37] See, e.g., Väyrynen, 2013.

[38] On mixed normative/social kinds, see Griffiths 2004.

[39] Gibbard (2003), Blackburn (1998).

[40] Sundell and Plunkett concede this point at Plunkett & Sundell, 2013a, pp. 28–30.

[41] Indeed, this paper itself can be seen as a form of metalinguistic negotiation over "political philosophy"/POLITICAL PHILOSOPHY.

[42] See Haslanger 2020.

REFERENCES

Benhabib, S. 1986. *Critique, Norm, and Utopia: A Study of the Foundations of Critical Theory*. New York: Columbia University Press.

Blackburn, S. 1998. *Ruling Passions: A Theory of Practical Reason*. Oxford: Oxford University Press.

Boyd, R. 1989. "What Realism Implies and What It Does Not." *Dialectica* 43: 5–29.

Boyd, R. 1991. "Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds." *Philosophical Studies* 61: 127–148.

Buchanan, A. 2013. *The Heart of Human Rights*. Oxford: Oxford University Press.

Coulthard, G. 2014. *Red Skin, White Masks: Rejecting the Colonial Politics of Recognition*. Minneapolis: University of Minnesota Press.

Cudd, A. 2006. *Analyzing Oppression*. Oxford: Oxford University Press.

Forrester, K. 2019. *In the Shadow of Justice*. Princeton: Princeton University Press.

Forst, R. 2017. *Normativity and Power: Analyzing Social Orders of Justification*. Oxford: Oxford University Press.

Geuss, R. 1981. *The Idea of a Critical Theory*. Cambridge: Cambridge University Press.

Geuss, R. 1999. *Morality, Culture, and History*. Cambridge: Cambridge University Press.

Gibbard, A. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Godman, M. 2020. *The Epistemology and Morality of Human Kinds*. London: Routledge.

Griffin, J. 2008. *On Human Rights*. Oxford: Oxford University Press.

Griffiths, P. 2004. "Emotions as Natural and Normative Kinds." *Philosophy of Science* 71: 901–911.

Hacking, I. 1999. *The Social Construction of What?* Cambridge, UK: Harvard University Press.

Hall, K. 2009. "Queer Breasted Experience." In *"You've Changed": Sex Reassignment and Personal Identity*, edited by Laurie Shrage, 121–135. Oxford: Oxford University Press.

Haraway, D. 1991. "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century." In *Simians, Cyborgs, and Women*, edited by Donna Haraway, 149–183. New York: Routledge.

Haslanger, S. 2000. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34: 31–55.

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

Haslanger, S. 2020. "Going on, Not in the Same Way." In *Conceptual Engineering and Conceptual Ethics*, edited by Herman Cappelen and David Plunkett, 480. Oxford: Oxford University Press.

Horkheimer, M. 1972. *Critical Theory: Selected Essays*, Trans. M. J. O'Connell. New York: Continuum.

Jaeggi, R. 2014. *Alienation*. New York: Columbia University Press.

Jenkins, K. 2016. "Amelioration and Inclusion: Gender Identity and the Concept of Woman." *Ethics* 126: 394–421.

Kaplan, D. 1989. "Demonstratives." In *Themes from Kaplan*, edited by Joseph Almog, John Perry, and Howard K. Wettstein, 481–563. Oxford: Oxford University Press.

Khalidi, M. A. 2015. "Three Kinds of Social Kinds." *Philosophy and Phenomenological Research* 90: 96–112.

Lawford-Smith, H. 2022. *Gender-Critical Feminism*. Oxford: Oxford University Press.

Le Guin, U. K. 1987. *The Left Hand of Darkness*. New York: Penguin.

Ludlow, P. 2008. "Cheap Contextualism." *Philosophical Issues* 18: 104–129.

MacCallum, G. C., Jr. 1967. "Negative and Positive Freedom." *The Philosophical Review* 76: 312–334.

Mason, R. 2016. "The Metaphysics of Social Kinds." *Philosophy Compass* 11: 841–850.

Plunkett, D., and T. Sundell. 2013a. "Disagreement and the Semantics of Normative and Evaluative Terms." *Philosopher's Imprint* 13: 1–37.

Plunkett, D., and T. Sundell. 2013b. "Dworkin's Interpretivism and the Pragmatics of Legal Disputes." *Legal Theory* 19: 242–281.
Sangiovanni, A. 2017. *Humanity without Dignity: Moral Equality, Respect, and Human Rights*. Cambridge: Harvard University Press.
Schroeter, L., and F. Schroeter. 2014. "Normative Concepts: A Connectedness Model." *Philosopher's Imprint* 14: 1–26.
Searle, J. 1995. *The Construction of Social Reality*. New York: Simon and Schuster.
Searle, J. 2010. *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press.
Shelby, T. 2009. *We Who Are Dark: The Philosophical Foundations of Black Solidarity*. Cambridge, MA: Harvard University Press.
Swift, A. 2008. "The Value of Philosophy in Nonideal Circumstances." *Social Theory and Practice* 34: 363–387.
Thomasson, A. 2017. "What Can We Do, When We Do Metaphysics?". In *The Cambridge Companion to Philosophical Methodology*, edited by Søren Overgaard and Giuseppina D'Oro, 101–21. Cambridge, UK: Cambridge University Press.
Thomasson, A. 2020. "A Pragmatic Method for Normative Conceptual Work." In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen and David Plunkett, 435–58. Oxford: Oxford University Press.
Valentini, L. 2012. "'Ideal Vs Non-Ideal Theory: A Conceptual Map." *Philosophy Compass* 7: 654–664.
Väyrynen, P. 2013. *The Lewd, the Rude and the Nasty: A Study of Thick Concepts in Ethics*. Oxford: Oxford University Press.
Williams, B. 2002. *Truth and Truthfulness*. Princeton: Princeton University Press.
Young, I. M. 1990. *Justice and the Politics of Difference*. Princeton: Princeton University Press.

AUTHOR BIOGRAPHY

Andrea Sangiovanni is Professor of Philosophy at King's College London. From 2018-2020, he was Professor of Social and Political Theory at the European University Institute, Fiesole. He is the author of Humanity without Dignity: Moral Equality, Respect, and Human Rights (Harvard University Press, 2017) and Solidarity: Nature, Grounds, and Value (forthcoming, Manchester University Press).

How to cite this article: Sangiovanni, Andrea. 2023. "Critical Theory, Ideal Theory, and Conceptual Engineering." *Journal of Social Philosophy* 1–18. https://doi.org/10.1111/josp.12545