[This is the penultimate draft. Please cite the final version of the paper, available at https://onlinelibrary-wiley-com.stanford.idm.oclc.org/doi/10.1111/phc3.12782]

Social structural explanation

Abstract: Social problems such as racism, sexism, and inequality are often cited as structural rather than individual in nature. What does it mean to invoke a social structural explanation, and how do such explanations relate to individualistic ones? This article explores recent philosophical debates concerning the nature and usages of social structural explanation. I distinguish between two central kinds of social structural explanation: those that are autonomous from psychology, and those that are not. This distinction will help clarify the explanatory power that each type of SSE has, points of convergence with methodological traditions such as critical theory and rational choice theory, and the difficulties that each type of SSE faces.

1. Introduction

What does it mean to say that structural racism explains persistent racial segregation, patriarchy explains ongoing gender inequalities, and capitalism explains widening income inequality? These are *social structural explanations* (SSEs): they appeal to some aspect of social structure, rather than to properties of individuals, to explain social phenomena. SSEs are typically contrasted with individualistic ones, such as: individual bias explains persistent racial segregation, sexism explains ongoing gender inequalities, and differences in work ethic explain widening income inequality.

SSEs intend to pinpoint fundamental causes of social problems, whereas individualistic explanations pick out only proximate causes. As a result, SSEs have gained traction in public discourse and social and political philosophy in recent years. But despite their prevalence, their purported advantages are underexplored, and some of their difficulties ignored.

The central claim of this paper is that we should distinguish between two forms of SSEs: those that are autonomous from psychology, and those that are not. Doing so will help clarify the explanatory power that SSEs have and the difficulties that they face. The paper proceeds as follows. Section 2 situates SSE as a specific form of methodological holism. Sections 3 and 4 clarify several distinct notions of SSE. Section 5 discusses the scope of SSEs and some issues with causally operationalizing them. Section 6 concludes with some notes on the connection between SSEs and “grand unified theorizing”.

2. Background: contemporary usage and holism

Contemporary usage

SSEs attempt to explain persistent social phenomena that seem unified into a pattern. For example, patterns of Black-white racial inequality persist in U.S. society. The level of household wealth that the median Black family possesses is far below that of the median white family; the

quality of public goods such as schools, environmental quality, and recreational facilities is much lower in segregated Black neighborhoods than in white ones; and a disturbingly high proportion of Black men are incarcerated. What explains these patterns of inequality? This is a question about the *maintaining* causes of racial inequality, rather than a question about its originating causes. Historians have extensively documented those originating causes, such as slavery, redlining, disinvestment in black communities (see especially Rothstein 2017), and individual-level racism. The question social theorists ask is about why these patterns persist even when the originating causes may have decreased in severity.

One common explanation for these patterns is *structural racism*. Such explanations refer to properties of social structure, such as norms, practices, and institutions, to explain both patterns of social phenomena, as well as individual cases that fit these patterns. Consider racial residential segregation as a paradigm example of structural racism. Residential segregation is caused and maintained, in large part, by federal housing policy, discriminatory lending policies such as redlining, and exclusionary residential zoning rules enacted by local jurisdictions (Denton & Massey 1993, Rothstein 2017, Trounstein 2018). These are formal elements of social structure: institutions and policies. Social structures also have informal elements such as social norms and practices, which are not written down anywhere. Norms are behavioral rules that are known to exist and apply to a class of situations (Bicchieri 2005), and we can think of practices as norm-following behavior (Haslanger 2016, 125-126). For example, norms about socializing with those who are dissimilar to us may explain why people tend to self-segregate by race in unstructured environments such as cafeterias (Schelling 1978, 137-166).

Methodological holism

Because SSE refers to properties of social structure, it is a form of *methodological holism*: the thesis that some social facts are best explained in terms of other social facts, rather than in terms of facts about individuals (see Zahle's (2016) overview). Holism's contrast class is methodological individualism, which holds that social facts are always (at least in principle) explainable in terms of lower-level facts about individuals and their interactions (see Heath's (2020) overview). This paper carves out SSE as a distinctive form of holistic explanation that focuses on forms of social organization rather than on general social facts. To illustrate the distinction, consider these two forms of explanation for the Black-white racial income gap:

> **General holist**: Differences in exposure to environmental pollutants and access to healthy foods explains health inequities between Blacks and whites.
>
> **Social structuralist**: Segregation, which divides Blacks and whites into different physical locations with different exposure to environmental pollutants and access to healthy food, explains health inequities between Blacks and whites.

The general holist simply points out a variety of social facts without pointing to a broader causal structure. By contrast, the structuralist pinpoints a form of social organization with a certain causal structure that *unifies* the cited social facts.

These are explanatory, not ontological, theses. Few holists claim that there is an ontologically independent social structure that exists "above and beyond" the interactions of individuals. Rather, contra the individualist (cf. Elster 1989, the claim is that a higher level of description provides more information than a lower one). Specifically, the holistic explanation provides more robust information – it explains why certain patterns hold across different contexts, regardless of the variation among the individuals that instantiate those patterns (List & Spiekermann 2013). SSE adds to general holism by appealing to a social structure to unify seemingly distinct events under a pattern (Haslanger 2016, 125). This structure is either causally responsible for the *content* of individual attitudes which then cause the problematic phenomenon, or it ensures that the problematic phenomenon would take place *regardless* of the content of individual attitudes. Either way, structure is the more fundamental cause of the problematic phenomenon.

The next two sections distinguish between these two notions of SSE, which I classify as non-autonomous and autonomous, respectively.

3. Non-autonomous explanation: the subversive model

Social structures shape our attitudes, beliefs, and values. Jackson & Pettit call this the "subversive model": structures subversively affect psychology, so we should point to structures rather than psychology as the causes of social phenomena (1992, 108-111). Psychology is merely the proximate cause. On this analogy, if a slate falls from a roof and causes a pedestrian to jump, thereby causing a car accident, we would point to the falling slate -- not to the pedestrian -- as the cause of the accident.

The subversive model is intuitively appealing because it makes sense of the observation that oftentimes, individuals' prejudiced mental states are a significant cause of social injustice. And it does not seem to be a mere coincidence that individuals are biased in the same ways, i.e. negatively against minorities. Thus, subversive modelers look toward social structures to explain why individual psychologies share these similarities: biased patterns of thought are shaped by structures such as institutions, norms, and material representations of social groups.

For example, Martín (2020) argues that "white ignorance," the fact that whites systematically underestimate the extent of racial inequality, and that this ignorance gives rise to racial domination, is best explained by structural processes that create ignorance of these facts. Anderson (2010) argues that even if racial segregation is in large part the result of racial stigmatization – a psychological process – these patterns of stigmatization can be traced to segregated environments themselves. Segregated environments generate racially biased attitudes. Thus, to ameliorate segregation that is due to racial stigmatization, we should promote integration via structural reforms in institutionalized settings, such as affirmative action.

The subversive model assumes that our social cognitive apparatus reliably tracks environmental regularities. Our exposure to certain material representations or descriptive norms, such as the fact that there are fewer women than men in STEM and more Blacks than whites in prison, creates semantic associations that track these regularities (Huebner 2016). But unjust social structures can cause our cognitive apparatus to issue in certain objectionable states such as

flawed perceptual skill, which is epistemically accurate but nevertheless biased because it tracks biased regularities (Munton 2019). So does our exposure to material objects. For example, Kodak's "Shirley cards" calibrated photo lighting for ivory skin. This normalized whiteness over other skin tones, creating light-skin bias in photography that partly constitutes racist bias (Liao & Huebner 2020, 3-5). As Liao & Huebner put it, "racist things shape racist thoughts. And they do so because they provide anchors for our attitudes, which lead us to backslide toward biases, even when we make good faith efforts to change our habituated patterns of thought and action" (Liao & Huebner 2020, 15). Thus, according to subversive theorists, explanations that refer to "bias" or "flawed perceptual skill" alone are inadequate. Such explanations blame biased minds solely on our cognitive apparatus, instead of also on the social environment that provides its input.

The subversive model underlies criticisms of implicit bias explanations for racial injustice. On this criticism, focusing on implicit bias is like playing whack-a-mole with racial injustice; bias is not the root cause of injustice. In Haslanger's words, "an adequate account of how implicit bias functions must situate it within a broader theory of social structures and structural injustice; changing structures is often a precondition for changing patterns of thought and action and is certainly required for durable change" (Haslanger 2015, 1).

Instead of implicit bias, critics appeal to social structural elements such as culture, especially norms and ideology. For example, Ayala-López (2018) explains discursive injustice in terms of conversational norms rather than speakers' biases. While the term "ideology" is notoriously vexed, critical theorists agree that it at least partly consists of beliefs and judgments that perpetuate unjust social relations by distorting social reality (Geuss 1981, Celikates 2006, Shelby 2003, 2017, Stanley 2015, Haslanger 2017). Shelby summarizes the point of critical theory thusly: "ideology-critique is indispensable for understanding and resisting the forms of oppression that are characteristic of the modern world" (Shelby 2003, 154). The subversive model points to criticism of ideology as necessary for social change.

Issues with the subversive model

The move towards ideology exposes the subversive model's reliance on a problematic ontological distinction between individuals and structure. To claim that structures are more basic causes than psychology, we must assume that there is a distinction between structures and psychology. But this distinction is fuzzy when it comes to ideology.

Ideology is, in part, a distorted cognitive schema (Haslanger 2017, 159-162). Shelby argues that ideology consists of a set of beliefs that have the following features: they are widely shared and known to be so; they form a coherent system with normative elements; they shape the self-conception of those in the relevant group; and they have a significant impact on social action and social practice (Shelby 2003, 158-159). And where are schemas or beliefs located but "in the head"? After all, to follow a norm, individuals reference (consciously or unconsciously) what they *ought* to do in particular contexts, and reference to the normative seems ineliminably mental (Bicchieri 2016, 65-66). Madva (2016), Davidson & Kelly (2018), and Soon (2020)

leverage this ontological fuzziness to argue that social phenomena are best explained by the interaction of individuals and structure, neither of which takes priority.

Secondly, the subversive model relies on an "oversimplified model of the mind": the "MIRROR view", which claims that our social biases are only effects, rather than also causes of, our social environment (Madva 2016, 705). This hypothesized one-way pipeline between the environment and the mind ignores the fact that our attentional goals render certain features of the environment more salient than others.

Thirdly, some structural factors do not seem to be the kind of things that can influence individual psychology. The general holist faces the issue of how a statistic like "a high rate of urbanization", or an abstract social fact like "stratification", can directly produce certain psychological states (Jackson & Pettit 1992, 110). The structuralist faces a similar problem. Can capitalism cause certain psychological states (e.g. Weber's Protestant work ethic, the profit motive, greed) that then perpetuate inequality? This seems difficult to establish. Capitalism describes a network of interrelated mechanisms, such as various forms of production and consumption, that together constitute a form of social organization. These finer-grained mechanisms directly interact with individual psychology; "capitalism" is a catch-all description for this system of market activity. Thus, there needs to be a mechanistic constraint on the kinds of structural factors that the subversive model can invoke (cf. Martín 2020, who prefers to leave possible mechanisms open).

These worries restrict the subversive model's scope. It works best where: individuals are clearly ontologically distinct from structures, structures influence individuals but not vice versa, and the structural factors invoked can plausibly influence psychology.

4. Autonomous explanation: the program model

A stronger form of SSE will be autonomous from psychology. In autonomous structural explanations,

> "the factor involved is meant to be explanatory under a variety of possible individual-level processes; its invocation does not point us to any particular psychological explanation…Many psychological possibilities remain open and so the account on offer, if it is truly explanatory, explains in abstraction from the particular individual-level processes that are at work; it is not just another way or referring us to a micro-explanation of the result explained" (Jackson & Pettit 1992, 103).

This is an argument from *multiple realizability*: the social phenomenon in question will arise whenever the structural cause is present, but that phenomenon is compatible with a variety of micro-level processes. Jackson & Pettit call this the "program model" of explanation (1992, 117-125). To illustrate, they offer this analogy. When a closed flask containing water is heated, it eventually cracks. What explains the cracking? Those who think that causal explanations should be maximally fine-grained (e.g. Elster 1989) will point to the causal chain at work in the actual

world, involving particular molecules colliding with one another at specific velocities. Call these *process* explanations.

Process explanations do not give us enough interesting information. We don't just want information about *this* particular flask, or about this particular social case. We seek generalizable information: Why is it that *whenever* a closed flask containing water is raised to boiling point, the flask cracks? Why is it that racial segregation, wherever it occurs, leads to worse outcomes for Blacks? Program explanations give us such information. The boiling point of water, and the fact that the flask is closed, are structural factors that "program for" the flask cracking. In other words, program explanations identify *conditions* that more or less ensure that the relevant event will occur (Jackson & Pettit 1992, 119). SSEs are best understood as program explanations. They give us information about event-types, not merely event-tokens.

With this methodological context in hand, we are now in a better position to understand what structural racism is, understood as an autonomous SSE. This term does not imply, suggest, or require that individually racist attitudes be widespread within a society. Neither is it shorthand for, or a higher-level description of, an aggregation of individually racist attitudes or actions within a society. Structural racism and other SSEs deliberately occlude reference to properties of individuals, such as their preferences and attitudes. Rather, the claim is that the causal structure of society itself maintains various patterns of racial inequality, so it is no surprise that certain patterns are robust across contexts and over time. Individual psychology is not important.

To make this concrete, consider this SSE for ongoing racial residential segregation: whiter communities support stringent land-use regulations that maintain homogeneity by pricing out people of color. Racist beliefs within white communities are not necessary to maintain segregation, as long as material incentives are in place for whites to support policies that have segregationist effects. As Trounstine puts it:

> "Once racist policies are in place, individual beliefs (e.g., racism) among individual beneficiaries of the system become largely irrelevant. Obviously, the level of racism among whites is both variable and impactful for political and economic outcomes...But, in the end, because government policy generates segregation through land use, the consequences of this variation are reduced. The choices of the racially resentful and the less racially resentful can become indistinguishable. Whites tend to make decisions that reinforce their privilege without thinking too deeply about it because they want stable property values, good schools, nice parks, and low-crime neighborhoods, and they have the financial opportunity to pursue those goals." (Troustine 2018, 208)

Here, the importance of racial ideology takes a backseat to incentive structures that organize behavior in ways that perpetuate injustice.

Rational choice and strategic interaction

One advantage of autonomous SSE over non-autonomous SSE is that the former can account for the role of strategic interaction in generating social phenomena. Táíwó (2018) criticizes subversive theories, especially Haslanger's and Stanley's, for neglecting strategic interaction, and argues that problematic social phenomena are best explained in terms of agenda-setting changes to incentive structures. Táíwó suggests that this framework provides guidance for empirical research.

A large literature in rational choice theory and economics, while not explicitly framed in terms of agenda-setting, has long modeled social phenomena as the product of strategic interaction in response to a dominant social agenda. We can think of rational choice theory (RCT) as a form of autonomous SSE that explains social phenomena, particularly oppression or structural injustice, as the result of individuals rationally and strategically maximizing their preferences under the constraints of social structures. As Young puts it, structural injustice "occurs as a consequence of many individuals and institutions acting to pursue their particular goals and interests, for the most part within the limits of accepted rules and norms" (Young 2010, 52).

The RCT model of SSE does not rely on ideology or other psychological states to explain why individuals and institutions act in ways that maintain injustice. As such, the RCT model has the advantage of explaining the self-maintaining nature of social phenomena without relying on an alignment between individual beliefs and the phenomenon in question. While a lack of criticism of these structures can, in part, be responsible for their persistence, RCT theorists emphasize that critical analysis is insufficient for social change. Even if individuals are aware that they are acting in response to perverse incentive structures and disagree with the ideology embodied by these structures, it is not instrumentally rational for them to act otherwise as long as sufficiently strong incentive structures remain. Thus, the problematic social phenomenon persists.

For example, people falsify their preferences in public because they perceive these preferences to be out of step with accepted views on controversial issues, such as affirmative action (Kuran 1997); practitioners of female genital mutilation do not approve of the practice in private, but feel the pressure of following the norm (Bicchieri 2005, 2017). As such, Sankaran charges that critical theorists face a dilemma: "Either their account of social change fails to account for important strategic impediments to social change, in which case it is inadequate, or it incorporates a theory of strategic behavior, and thus merely reinvents the wheel, poorly" (Sankaran 2020, 1441). Ideology should be understood non-ideationally in terms of conventions, or equilibrium solutions to social coordination problems. This understanding of ideology can be reconciled with RCT. Ideology, understood in the conventional, autonomous sense, incentivizes individuals to act in ways that align with that ideology, understood in the ideational, non-autonomous sense, even if they do not believe in that ideology.

Consider gender norms. Patriarchal gender norms continue to burden women with the lion's share of domestic labor (Miller 2020). This is a leading cause of the gender wage gap; men are able to move into higher positions in the workforce compared to women with family responsibilities (Goldin et al 2017). Some of these women and their husbands may buy into patriarchal ideology and believe that women ought to take on the bulk of domestic labor. But they need not hold these beliefs to make the same choice. Lisa, a feminist who doesn't like

domestic labor, might nevertheless quit her job because her husband makes more money than she does and childcare is expensive, so the most economical decision for her family is for her to stay home (Okin 1989, Cudd 2006, Haslanger 2016).

The key point is that we cannot read off individuals' beliefs and values from their choices. Given certain norms, a variety of beliefs are compatible with a given choice. To explain these choices, we have to look at the social structures that make some choices more viable or attractive than others.

The RCT framework's emphasis on strategic behavior can explain why individuals *rationally* act in ways that seem to maintain their oppression. While subversive theorists posit the psychological kind of ideology to explain such behavior, this approach is criticized for attributing irrationality or Marxian false consciousness to agents. RCT avoids these criticisms by explaining how "individuals often get outcomes they do not want, not because they have chosen wrongly, but because they have chosen instrumentally" (Heath 2000, 365). Cudd's (2005, 2006) structural RCT articulates this framework thusly: it "explains the maintenance of oppression as the maintenance of unjust social conventions and social practices by individuals who are motivated by those conventions and practices to act to maintain them" (Cudd 2005, 27). Thus, strategic explanation satisfies one normative criterion that ideological explanation does not: it respectfully models the oppressed as rational agents who are responding to the constraints of their situation, rather than people who are systematically misled about what is in their best interests (Khader 2012). For example, we can explain the persistence of beauty standards as a result of the fact that beauty is a positional good. It is not enough to simply be attractive; one has to be more attractive than others to be considered beautiful. Because beauty comes with significant social and material advantages, women have a strong incentive to move up in the beauty hierarchy even if they disavow patriarchal ideology (Heath 2000, 369).

The RCT or strategic model has relevance to the longstanding debate in feminist theory concerning why women in the global South participate in traditions that apparently maintain their oppression, e.g. the Islamic practice of wearing the burqa. On what Khader calls Western "missionary feminism", "other" women need to be saved *from* their own cultures and saved *to* Western culture; this "saving to" crucially includes the inculcation of Enlightenment values such as individual autonomy (Khader 2018, 22-49). Khader argues that a more empirically accurate, less ideologically blinkered feminism would take into account the fact that in non-Western cultural contexts, women genuinely secure their well-being by participating in forms of life that are not tied to the autonomy ideal. They act strategically to secure their well-being, and have heterogeneous reasons for doing so. For example, in a pastoral society where economic opportunities tied to education are absent, women act in their best interest by marrying into the right family rather than pursuing education (Khader 2018, 65-66). This explanatory perspective highlights the nature of the problem with patriarchal practices: their bundling with goods such as social status constitutes a structural constraint that individual women cannot overcome (Khader 2018, 71-72). Hence, a more productive feminist strategy for change ought to focus on unbundling goods and practices rather than on changing ideology.

At this point, we might encounter Haslanger's criticism that RCT is too individualistic, because it explains behavior in terms of psychological states (Haslanger 2016, 121-123). That is, RCT explains behavior in terms of individuals strategically acting to maximize their preferences, and "preference" is a psychological notion. In light of this criticism, one might wonder why RCT aligns with SSE rather than individualism. Setting aside the debate about the nature of preferences in economic theory (cf. Rosenberg 1992, Hausman 1992), recall the previous point about multiple realizability: if a variety of psychological states realize a social phenomenon, then psychological states are not explanatory. This does not mean that psychological states are ontologically unnecessary; some psychological states (e.g. beliefs, desires) are necessary for any RCT explanation to work. Rather, to say that psychological states are not explanatory is to say that the individual's action does not depend on them having a particular psychological state. Given a constraining enough structure, individuals will act in the same way even if they have heterogenous preferences and reasons (Satz & Ferejohn 1994). Thus, although Haslanger distances herself from RCT, in my view, her theory is also a structural RCT. For Haslanger, the social positions that individuals occupy within structures best explain their choices; Lisa rationally makes the decision to quit her job because she is a woman who faces a certain set of incentive structures that men do not, in virtue of her social position: "women as a group are structurally situated so that it is rational for them to choose options that keep them subordinate" (Haslanger 2016, 124).

While I have laid out a conceptual distinction between autonomous and non-autonomous SSE, in reality, both forms of SSE are likely useful for explaining many social phenomena. Often, incentive structures influence psychology; what's descriptively normal becomes prescriptively normal over time. Racist attitudes can develop and eventually sustain segregated environments, and sexist attitudes can develop under, and sustain, oppressive gender norms. The point made by proponents of autonomous SSE is simply that we cannot infer psychology from behavior, as a heterogenous set of reasons or mental states are compatible with a behavior. As Khader puts it: "That an agent in a context with sexist norms eats less than her husband does not yet tell us whether she believes she is a lesser human being, seeks favour with her male relatives, wants to feel like a 'dutiful woman' - or something else entirely" (Khader 2012, 313). Distinguishing between these two types of SSE helps us explore more possibilities for modeling behavior, instead of directly inferring psychology from behavioral patterns.

5. Varieties of explanation

Up to this point, we have laid out a framework for understanding what makes a social explanation *structural*. Now to a deeper question: what kind of *explanation* are SSEs? Here I address the scope of SSEs and some challenges.

Explanatory scope and erotetic explanation

Many (though not all) autonomous SSEs are ambitious in scope. While some only attempt to explain a particular phenomenon (Heath 2000), others attempt to explain the persistence of these phenomena in society at large (Cudd 2005, 2006, Haslanger 2016). This is an explanatory demand: "We need something unifying to explain the striking commonality of

gender oppression" (Barnes 2017, 2424), because the commonality of oppression does not seem like a mere coincidence. A disunified explanation may even lead us astray by understating the scope of the problem. As Frye (1983) writes:

> "The experience of oppressed people is that the living of one's life is confined and shaped by forces and barriers which are not accidental or occasional and hence avoidable, but are systematically related to each other in such a way as to catch one between and among them and restrict or penalize motion in any direction." (Frye 1983, 3)

SSEs thus aim to subsume seemingly unrelated, but similar, events under a unified explanation (Kitcher 1981). An explanation in terms of patriarchal social structures provides a unified explanation of the gender wage gap, unequal domestic labor burdens, and women's financial reliance on their male partners. Together, these individual constraints constitute an interlocking system of constraint: oppression (Frye 1983, Young 2014). These general explanations apply more broadly and are therefore more stable, in the sense that they are insensitive to individual variations (Haslanger 2016, 119). So, independent, disunified explanations are unsatisfactory, even if technically causally correct.

SSEs are thus best understood as erotetic explanations: they are answers to why-questions, and questions establish the contrast class for the appropriate answer (van Fraassen 1980, Garfinkel 1981, Cross 1991). Why does Lisa, rather than her husband, quit her job? The best answer to this question should refer to the social structure in which Lisa is embedded, because structure constrains the possibilities available to her (Haslanger 2016, 114-118). This answer draws attention to the context in which Lisa's behavior takes place, and explains the behavior of similarly situated women: "The explanations of the workings of the structure will be the best way to explain the workings of its parts" (Haslanger 2016, 118).

Broad and deep vs. local and flexible explanations

The idea of social structure is still vague; there are different scales of social structure. Which is most explanatorily edifying -- "broad and deep" structures such as patriarchy and capitalism, or "local and flexible" institution-specific structures such as churches or businesses (Haslanger 2016, 113)? While Haslanger defends the importance of broad and deep structural explanations, Sterken (2018) argues that local and flexible structural explanations of individual decisions are sometimes preferable to broad and deep ones. Local structures are ones that are tied to a particular context, and flexible structures are ones that are less modally robust (Sterken 2018, 185). Whereas Haslanger would characterize Lisa's situation in terms of high-level structural generalizations such as the wife-mother relation, or the employer-employee relation, Sterken thinks it is more fruitful to explain Lisa's situation in terms of local, contextual structures, such as whether the employer is good or crappy with respect to parental leave, because those are most immediately relevant to Lisa's decision.

This challenge to broad and deep structures is not a general objection to structural rational choice SSE, which is flexible enough to apply to more or less fine-grained contexts. It only

imposes a constraint on the scope of such an explanation. In response to both Sterken and Haslanger, we might think of broad social structures as high-level generalizations about features of local-level structures. Gender norms are considered *broad* social structures because they pervade so many local contexts in roughly the same ways. Broad SSEs are useful because they capture the cross-contextual similarities of these norms, but their high level of generality may sometimes lead to explanatory inadequacy.

Causal explanation as inherently normative

One major point of contention in this debate concerns the selection of causes. The contrast class with structural explanations is individual-level explanations, but both individuals and structures seem ontologically necessary for any social phenomenon to take place. Why foreground structures in social explanation? This question touches on the distinction between enabling conditions and causes: in any causal structure, some factors are "backgrounded" as enabling conditions and others are selected as causes. For example, the disaster wrought by a hurricane is both a function of inadequate infrastructure and the severe weather event itself. It is common to say that the natural disaster is caused by the hurricane and relegate inadequate infrastructure to an enabling background condition. But as Lewis argues, this selection is causally arbitrary (Lewis 1973, 559) – no causal principles can make that distinction for us.

So, why take capitalism to be *the* cause of poverty rather than individuals' choices within that system, given that both are jointly necessary for poverty to occur? (Zheng 2018, 340) Zheng adopts Kronfeldner's (2014) view of causal selection to answer this question: "we select as causes the factors that we are willing and prepared to change" (Zheng 2018, 330). Our normative commitments influence whether we take a causal factor as fixed (an enabling condition) or changeable (the cause). Zheng argues that moral philosophers have an important role to play in determining causal explanations; they can excavate the moral commitments underlying causal selection and thereby push for moral progress.

This view might bring to mind Nozick's quip: "Normative sociology, the study of what the causes of problems ought to be, greatly fascinates us all" (1974, 247). Nozick's offhand comment is directed at what he perceives to be a tendency to select causal explanations that are congenial to our political views, at the expense of dismissing the actual causes of problems. Is Zheng's view vulnerable to this criticism? No. For Zheng and others, causal structure is objectively fixed; the inherent normativity of causal selection concerns our emphasis on various components of this objective structure. But this view is subject to difficulties that have to do with operationalizing structure in causal language. The next section focuses on these difficulties.

Causation as intervention

On Woodward's (2003) popular interventionist account of causal explanation, X is a cause of Y if and only if an intervention on X would make a difference to Y without changing any other variables in the system. This underlying notion of explanation is attractive because it explicitly bridges the gap from explanation to intervention, satisfying a pragmatic aim (Bright et al 2016, 76). But underwriting SSE with the interventionist account raises some issues.

The trouble is that capitalism, white supremacy, or patriarchy are not independently manipulable variables; they are “macro-level” features about the causal structure of a system (Malinsky 2018). Each of these structures describes a set of norms, practices, attitudes, and material features that are interrelated in complex ways. Will transitioning from capitalism to socialism reduce poverty? Will a socialist society be less structurally racist than a capitalist one? It is difficult for interventionists to answer this question because even if we represent these features as variables, they are probably not possible to manipulate independently of other variables: “Causal claims about social features including patriarchy and capitalism are confusing because it does not seem like such interventions are possible, even in principle” (Malinsky 2018, 2298).

Steel calls these “structure-altering interventions”: “an intervention on X is *structure altering with respect to* V just in case it changes causal relationships among the variables of V *in addition to* eliminating the causes of X" (2006, 450). For Steel, structure-altering interventions are only uncogently used in causal explanation. Malinsky, however, thinks we can make sense of structural claims in causal explanation. Malinsky suggests that we identify structural claims with sets of causal parameters in structural equation models, and “interpret counterfactuals about structural features as claims about alternative parameter settings in these causal models” (2018, 2296). This operationalization of social structural claims allows us to investigate them using machine-implementable statistical methods. More work is needed to refine a theory of SSE using this operationalization.

6. Conclusion: intervention and the perils of grand unified theorizing

This review has distinguished between two forms of SSE: the subversive model, which is not autonomous from psychology, and autonomous explanations, which emphasize strategic behavior under constraint. The latter are especially interesting because their power lies in their unifying ability: they explain how independent events are related by reference to a singular explanation -- social structure.

I want to end with some notes of caution about the use of SSEs. While SSEs’ signal strength is their ability to unify events under a pattern, this strength also leaves SSE vulnerable to the complaint that they are “Grand Unified Theories of Social Structure and Change” (Madva 2019). Madva argues that these make two related epistemic errors.

First, they assemble independent events into a pattern and then seek an explanation for the pattern. While it seems intuitive that social disparities relating to race, gender, and other social categories are related, we should be wary of overgeneralizing about these connections. As Madva puts it, the danger is that “The unchecked impulse to subsume a wide and variegated range of phenomena under a simplistic theoretical roof can generate distortions, omissions, and post hoc rationalisations of unruly data points that do not fit easily into the picture” (Madva 2019, 10). Second, assembling events into one explanandum creates the impulse to seek foundational, “linchpin” explanations. This impulse can obscure the operation of different kinds of mechanisms that cause the phenomena of interest. Distinguishing between autonomous and non-autonomous forms of SSE is one step toward roughly differentiating between two types of mechanisms.

Harris (2018) has similar worries about attempts to unify instances of anti-Black racism under a coherent structural explanation. The diversity of Black experience across the globe undermines the explanatory power of any structural explanation: “to explain their worlds under one rubric arguably requires an unlimited number of caveats” to take into account immigration status, ethnicity, language, and other factors (Harris 2018, 287).

In spite of these cautionary notes, social theory requires unified theorizing, which will often exceed the bounds of empirical evidence. Future work should focus on integrating a causally operationalizable framework for structural claims with social theory and empirical evidence.

References

Anderson, Elizabeth. (2010). *The Imperative of Integration*. Princeton University Press.

Ayala-López, S. (2018). A Structural Explanation of Injustice in Conversations: It’s About Norms. *Pacific Philosophical Quarterly*, *99*, 726–748.

Barnes, E. (2017). Realism and Social Structure. *Philosophical Studies*, *174*, 2417–2433.

Bicchieri, C. (2005). *The Grammar of Society*. Cambridge University Press.

Bicchieri, C. (2016). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press.

Bright, L. K., Malinsky, D., & Thompson, M. (2016). Causally Interpreting Intersectionality Theory. *Philosophy of Science*, *83*, 60–81.

Celikates, R. (2006). From Critical Social Theory to a Social Theory of Critique: On the Critique of Ideology After the Pragmatic Turn. *Constellations*, *13*(1), 21–40.

Cross, C. B. (1991). Explanation and the Theory of Questions. *Erkenntnis*, *34*, 237–260.

Cudd, A. E. (2005). How to Explain Oppression: Criteria of Adequacy for Normative Theories. *Philosophy of the Social Sciences*, *35*(1), 20–49.

Cudd, A. E. (2006). *Analyzing Oppression*. Oxford University Press.

Davidson, L. J., & Kelly, D. (2020). Minding the Gap: Bias, Soft Structures, and the Double Life of Social Norms. *Journal of Applied Philosophy*, *37*(2), 190–210.

Frye, M. (1983). *Oppression and the Use of Definition*.

Garfinkel, A. (1981). *Forms of Explanation: Rethinking the Questions in Social Theory*. Yale University Press.

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

Glaeser, E. (2011). *Triumph of the City*. Penguin.

Goldin, C., Kerr, S. P., Olivetti, C., & Barth, E. (2017). The Expanded Gender Earnings Gap: Evidence from the LEHD-2000 Census. *American Economic Review*, *107*(5), 110–114.

Harris, L. (2018). Necro-Being: An Actuarial Account of Racism. *Res Philosophica*, *95*(2), 273–302.

Haslanger, S. (2015). Distinguished Lecture: Social Structure, Narrative, and Explanation. *Canadian Journal of Philosophy 45*(1), 1-15.

Haslanger, S. (2016). What is a (Social) Structural Explanation? *Philosophical Studies*, *173*, 113–130.

Haslanger, S. (2017). Culture and Critique. *Aristotelian Society Supplementary Volume*, *91*(1), 149–173.

Hausman, D. M. (1992). *The Inexact and Separate Science of Economics*. Cambridge University Press.

Heath, J. (2000). Ideology, Irrationality, and Collectively Self-Defeating Behavior. *Constellations*, *7*(3), 363–371.

Heath, J. (2020). Methodological Individualism. In Edward. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. URL=<https://plato.stanford.edu/archives/sum2020/entries/methodological-individualism/>

Huebner, B. (2016). Implicit Bias, Reinforcement Learning, and Scaffolded Moral Cognition. In *Implicit Bias and Philosophy, Vol.1: Metaphysics and Epistemology* (pp. 47–79). Oxford University Press.

Jackson, F., & Pettit, P. (1990). Program Explanation: A General Perspective. *Analysis*, *50*(2), 107–117.

Jackson, F., & Pettit, P. (1992). Structural Explanation in Social Theory. In D. Charles & K. Lennon (Eds.), *Reduction, Explanation and Realism*. Oxford University Press.

Khader, S. J. (2012). Must Theorizing About Adaptive Preferences Deny Women's Agency? *Journal of Applied Philosophy*, *29*(4), 302-317.

Khader, S. J. (2018). *Decolonizing Universalism: A Transnational Feminist Ethic.* Oxford University Press.

Kitcher, P. (1981). Explanatory Unification. *Philosophy of Science*, *48*(4), 507–531.

Kronfelder, M. (2014). Commentary: How Norms Make Causes. *International Journal of Epidemiology*, *43*(6), 1707–1713.

Kuran, T. (1997). *Private Truths, Public Lies*. Harvard University Press.

Lewis, D. (1973). Causation. *Journal of Philosophy*, *70*(17), 556–567.

Liao, S., & Huebner, B. (2020). Oppressive Things. *Philosophy and Phenomenological Research*.

List, C., & Spiekermann, K. (2013). Methodological Individualism and Holism: A Reconciliation. *American Political Science Review*, 629–643.

Madva, A. (2016). A Plea for Anti-Anti-Individualism: How Oversimple Psychology Misleads Social Policy. *Ergo*, *3*(27).

Madva, A. (2019). Integration, Community, and the Medical Model of Social Injustice. *Journal of Applied Philosophy*.

Malinsky, D. (2018). Intervening on Structure. *Synthese*, *195*, 2295–2312.

MartÍn, A. (2020). What Is White Ignorance? *The Philosophical Quarterly*.

Massey, D., & Denton, N. A. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press.

Miller, C. C. (2020, February 11). Young Men Embrace Gender Equality, But They Still Don't Vacuum. *New York Times*. https://www.nytimes.com/2020/02/11/upshot/gender-roles-housework.html

Munton, J. (2019). Perceptual Skill and Social Structure. *Philosophy and Phenomenological Research*, *99*(1), 131–161.

Nozick, R. (1974). *Anarchy, State, and Utopia*. Basic Books.

Okin, S. (1989). *Justice, Gender, and the Family*. Basic Books.

Rosenberg, A. (1992). *Economics—Mathematical Politics or Science of Diminishing Returns?* University of Chicago.

Rothstein, R. (2017). *The Color of Law: A Forgotten History of How Our Government Segregated America*. Liveright.

Sankaran, K. (2020). What's New in the New Ideology Critique? *Philosophical Studies*, *177*(5), 1441–1462.

Satz, D., & Ferejohn, J. (1994). Rational Choice and Social Theory. *The Journal of Philosophy*, *91*(2), 71–87.

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. W. W. Norton.

Shelby, T. (2003). Ideology, Racism, and Critical Social Theory. *The Philosophical Forum*, *34*(2), 153–188.

Stanley, J. (2015). *How Propaganda Works*. Princeton University Press.

Steel, D. (2006). Methodological Individualism, Explanation, and Invariance. *Philosophy of the Social Sciences*, *36*(4), 440–463.

Sterken, R. K. (2018). The Structures of (Social) Structural Explanation: Comments on Haslanger's What is (Social) Structural Explanation? *Disputatio*, *X*(50), 173–199.

Táíwó, O. O. (2018). The Empire Has No Clothes. *Disputatio*, *51*, 305–330.

Trounstine, J. (2018). *Segregation by Design: Local Politics and Inequality in American Cities*. Cambridge University Press.

van Fraassen, B. C. (1980). *The Scientific Image*. Clarendon.

Weber, E., & Van Bouwel, J. (2002). Symposium on Explanations and Social Ontology 3: Can We Dispense with Structural Explanations of Social Facts? *Economics and Philosophy*, *18*, 259–275.

Woodward, J. (2003). *Making Things Happen*. Oxford University Press.

Young, I. M. (2002). Five Faces of Oppression. In J. E. Kelly (Ed.), *Industrial Relations* (pp. 174–202). Routledge.

Young, I. M. (2010). *Responsibility for Justice*. Oxford University Press.

Zahle, J. (2016). Methodological Holism in the Social Sciences. *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2016/entries/holism-social/>.

Zheng, R. (2018). A Job for Philosophers: Causality, Responsibility, and Explaining Social Inequality. *Dialogue*, *57*, 323–351.