Abstract: Feminists debate whether women can autonomously embrace their own subordination. Some argue that it is the process of identifying with desires and values that matters; others, that it is the content of the desires and values that matters. In this paper, I introduce a novel class of cases of ‘thwarted autonomy,’ in which women pursue autonomy but in ways that reinforce gendered subordination, and draw on these cases to develop an expressivist view of women’s autonomy. On this view, agents must embody desires and values in the social world if they are to achieve self-understanding, such that the social world mediates the relation-to-self. Oppressive meanings and norms can channel this expressive activity in ways that violate women’s physical integrity and involve self-abnegation that is incompatible with autonomy.

“An individual cannot know what he is until he has made himself a reality through action.”
- G. W. F. Hegel

“She chooses to want her enslavement so ardently it will seem to her to be the expression of her freedom.”
- Simone de Beauvoir

1. Introduction

To feminist philosophers, the ideal of autonomy presents both a promise and a challenge. On standard views, autonomy consists not in the absence of constraint, but in the ability to reflectively endorse the desires, ends, and values that underlie one’s action as ‘one’s own.’ Gendered socialization, feminists have long argued, shapes women’s desires to align with patriarchal demands. This naturalizes women’s oppression and, as women’s willing participation in sexist practices may be thought to immunize such practices from criticism, makes it hard to critique. As the concept of autonomy shines a critical light on an agent’s relationship to her own desires as the site of self-determination or lack thereof, it can articulate how oppression undermines women’s autonomy, even absent external constraints. To reap this concept’s critical promise, however, it is necessary to transform a traditional,
individualistic conception of autonomy into a robustly social one. Feminist philosophers have taken up this philosophical task: reconceptualizing autonomy in a way that captures (1) the socially constituted nature of the self - the way agents are embedded in social contexts and relationships - and (2) how an oppressive social context, in particular, can undermine women’s autonomy.

It is widely agreed that autonomy must be understood in a more social register. But the question of what it means to do so is controversial. The debate often begins with disagreement about so-called ‘hard cases’ of women who embrace gendered subordination. Thomas Hill’s ‘Deferential Wife,’ for instance, happily plays a subservient role in relation to her husband, and “tends not to form her own interests, values, and ideals; and, when she does, counts them as less important than her husband’s” (Hill 1973: 89). Other oft-discussed cases include Marina Oshana’s ‘Taliban Woman’ who chooses a life of extreme dependence out of religious piety (Oshana 2006; Westlund 2009), women who get plastic surgery (Chambers, 2008), or who accept gender-based injustices like lower wages without complaint (Nussbaum 2001). Some interpret these cases as ones in which oppression has insidiously compromised women’s autonomy. They seek to capture this intuition with substantivist views, on which autonomy is not content-neutral. Strong-substantivists place direct constraints on the content of women’s preferences or relationships (Stoljar 2000: 94-111; Oshana 2006; Hill 1973), while weak-substantivists build ‘content’ into their accounts indirectly: autonomy may require certain normative competencies, or traits like self-respect and imagination (Benson 1990, 1991, 1994; Wolf 1987; MacKenzie 2008). Proceduralists, however, worry that content-laden views of autonomy raise difficulties: they risk imposing a conception of the good on agents, rule out women who embrace traditional roles as non-autonomous, and can justify coercion or disrespect. They argue that autonomy must be understood as content-neutral (Christman 1990, 2001, 2004; Friedman 1997, 2003; Meyers 1987, 1989), such that autonomy does not depend on the content of a woman’s desires, but

---

4 For critical appraisals of the specifically masculinist bent of traditional conceptions of autonomy, see Code (1991) and Jaggar (1983).

5 Meyers (2000) moves away from a purely content-neutral account. For discussion of Meyers’s view, see Benson (2005).
on the way in which she comes to identify with them. Women can, on this approach, autonomously choose gendered subordination.⁶

Questions of social ontology about the nature of the social self underlie this debate. Yet these questions have remained implicit.⁷ This article brings this philosophical terrain to light in order to motivate an expressivist view of the social self as a promising foundation for an account of women’s autonomy. To shift the debate onto new conceptual ground, I focus on a pervasive but less discussed set of cases I term ‘thwarted autonomy,’ in which women value and pursue autonomy, but in ways that reinforce their own gendered subordination. By revealing how an oppressive social context can channel women’s attempts to actualize their autonomy in ways that come to undermine it, these cases highlight the need for a novel conception of the social self to underpin an account of autonomy. Drawing on Hegelian ideas of agency, I argue that the process of self-understanding and self-discovery that often constitutes autonomy’s core requires agents to first express their initially inchoate desires and values in a social context. Social meanings and norms, and the practices in which they inhere, mediate this expressive activity, and the way an agent interprets herself and her desires. On this framework, cases of thwarted autonomy can appear as the result of the limited, problematic vehicles available to women, through which they can express a desire for autonomy. This article motivates and explicates this view of the social self, and show how it illuminates the cases. While a full account of an expressivist view of women's autonomy must await future elaboration, I describe some of its features in the final section.

⁶ In the closely related debate about adaptive preferences, meanwhile, theorists disagree about how to interpret women's outward conformity to patriarchal norms in the first place, and the ethics of coercive intervention. While this closely overlaps with that over women's autonomy, there are broad differences in focus. The former originates in the framework of rational choice theory, and often focuses on the nature of adaptive preferences themselves, and practical questions they raise concerning intervention into people's lives. The aspect of this debate most germane to my purposes is the debate over how to interpret women's apparent acquiescence to patriarchal practices, and I discuss Uma Narayan’s intervention into this debate in §1. For other discussions of adaptive preferences, in addition to Nussbaum (2001), see Terlazzo (2016), Superson (2003), Elster (1983), and Cudd (2006).

⁷ The issue of social ontology appears in the distinction between causal and constitutive views of autonomy; often - though not always - this tracks the distinction between proceduralist and substantivist views. As I seek to emphasize the way in which social ontology matters to this debate, I will draw out underlying images of the social self accounts of autonomy presuppose. This means that in discussing substantivist and relational views, in §1 and §3 respectively, I group together views that diverge on questions of content, but converge on questions of social ontology.
The structure of the argument, in brief, is that if autonomy is to be conceptualized so as to capture (1) the socially constituted nature of the self and (2) the way oppression can undermine women's autonomy, an expressivist account furnishes an understanding of (1) in order to capture an under-theorized phenomenon that falls into the category of (2). The article is organized into three sections. In §1, I motivate the expressivist view by describing and drawing on two cases of thwarted autonomy to reveal the limitations of proceduralist and substantivist views. As proceduralists rely on a causal model of the social self, they fail to bring into view the way an oppressive culture can undermine autonomy. Substantivist views are equipped to capture this, meanwhile, but their emphasis on internalized norms renders them unable to account for the ambiguity of thwarted autonomy: the fact that women pursue the right value. §2 sketches the expressivist view's key insight: that agents gain self-understanding through the activity of expressing their initially unformed desires and values in a social context. The expressivist view borrows the proceduralist insight that autonomy is an activity that yields self-understanding, but recasts it as the practical, reflective articulation of desires and values in the world. It also borrows the substantivist insight that ‘content’ matters, but emphasizes the way a social context will partially construct this expressive activity. In §3 I show how this framework accounts for the dynamics of the cases from §1. I also consider what an expressivist view of autonomy entails, one crucial implication of which is that a social conception of autonomy must expand beyond the interpersonal realm to include collective hermeneutic resources and the normative structure of social practices.

I. Thwarted Autonomy

To motivate the expressivist view, I begin by considering the phenomenon of thwarted autonomy and the challenge it poses to proceduralist and substantivist approaches. Cases of thwarted autonomy are ones in which women value and pursue autonomy, but in ways that reinforce their own subordination. A case in point is that of Miriam Weeks who, as a student at Duke University in 2014, was revealed by a male classmate to have

On relational approaches, relationships to others are constitutive of autonomy. These views differ in whether they take autonomy to be content-neutral or content-laden. Oshana (2006) argues for a strongly-substantivist relational account, MacKenzie (2008), for a weakly substantivist relational account, and Westlund (2009), for a proceduralist relational account. I discuss MacKenzie and Westlund’s views in §3.
participated in physically abusive, misogynistic pornography. The media narrative seized not only on Weeks’s status as a student at an elite university, but also on the purportedly feminist motivations she elaborated in interviews. As she told one interviewer, participating in the porn industry was not only her choice, but one that made her feel like a “strong, independent woman” (Van 2014). As she explained, “everything is on my terms. I can say no whenever I want to. I am in control.” Even more directly, she stated that “[w]hat porn has done for me is it has given me back my agency” (Van 2014). Yet these claims were hard to square with the misogynistic content of the pornography, in which she was filmed being sexually degraded, verbally abused, and crying while being choked. Weeks herself later expressed regret over her participation in the industry, describing the experience as exploitative. Feminists, meanwhile, were split between those who saw Weeks as a victim of a predatory industry, and those who saw her as a woman shamed for publicly expressing her sexuality.

This case shares some features with the more familiar ‘hard cases.’ Weeks is not obviously coerced into participating in pornography, but does so of her own volition. As she can articulate her reasons for doing so, her decision does not seem entirely unreflective. If autonomy consists in an agent’s reflective identification with her desires and values, Weeks can be autonomous: indeed, she may exemplify the possibility that women can autonomously choose gendered subordination. Yet, as Stoljar has argued, the ease with which standard views can reach this conclusion raises the ‘feminist intuition’ (Stoljar 2000: 95) that preferences that have been shaped by oppression cannot be autonomous. This case both raises and complicates this intuition. As Weeks’s choice mirrors oppressive norms that encourage sexual submissiveness in women the case does, indeed, raise the feminist intuition that her autonomy has been undermined by her social context. But while women in ‘hard cases’ are often described as having a preference for subordination, this description is not quite apt. For in describing her motivations, Weeks cites a cluster of values diametrically opposed to subordination, such as strength, agency, control, and independence -
values that, taken together, look a lot like an ideal of autonomy. While this does not mean there is no sense in which Weeks may be said to prefer subordination, it does indicate the need for a view that can capture the way oppression can compromise women’s autonomy, even as they pursue it as a central value.

A second case will, shortly, furnish a more detailed example of this puzzling dynamic. But first, let us consider how a proceduralist approach fares in capturing the phenomenon. Proceduralist views are content-neutral: they do not specify what an agent must desire to be autonomous, but attend instead to the way in which an agent comes to identify with her desires. There is a diversity of proceduralist views but, generally, they take autonomy to have two conditions. First, agents must engage in a process of critical self-reflection, in a way that is conducive to self-awareness and self-appraisal: an agent may, for instance, have to display minimal rationality, and be free from delusion, paranoia, and self-deception. Second, this process must yield a form of identification with desires and values, whether understood in a minimal sense as a failure to be alienated from them (Christman 2004: 153), or in a stronger sense, as a capacity to affirm them as deeply significant (Friedman 2003: 14). By virtue of the proceduralist commitment to content-neutrality, women who meet these two conditions can, in principle, choose gendered subordination autonomously.

---

9 I appeal to women’s first-personal reports to characterize what their action means, in both cases of thwarted autonomy. One might argue that we ought not to take their first-personal perspectives seriously, as the women may be lying or deluded. While I do not take the women’s perspectives as the final world on what their action means, I do take them to be an important starting point in understanding the significance of their action: even if there are defects in the way Weeks understands her own activity, it provides important insights into the nature of the activity itself.
The socio-ontological foundations of the proceduralist approach, I argue, make it ill-equipped to identify the full range of ways oppression undermines autonomy. While traditional proceduralist views can be criticized for relying on a notion of a ‘true self,’ purged of social influence, the charge does not stick against feminist proceduralists. Rather than deny that the self is social, they argue that a proceduralist approach is perfectly compatible with this claim and, by extension, the claim that autonomy has social conditions. Still, this approach’s commitment to content-neutrality places limits on the way it can render this idea intelligible. It is useful to distinguish here between the claim that social conditions have a causal effect on autonomy, and the claim that they are constitutive of autonomy. On the former, egalitarian relationships - to take just one example - impact the extent to which agents can develop the skills they need for autonomy. Social conditions are “background requirements for the development of autonomy” (Christman 2004: 158), or “causal conditions” (Friedman, 2003: 14) for the development of autonomy-enhancing skills like understanding one’s desires, acting on them appropriately, and reflecting on alternatives. On the latter, constitutive version of the claim, egalitarian relationships can define autonomy, such that one could not be autonomous in their absence. Proceduralists largely conceptualize the relation between autonomy and social conditions as causal because a constitutive model risks violating their core commitment to content-neutrality. If one makes egalitarian

---

10 This argument may be framed by a meta-theoretical point. A curious feature of the debate over women’s autonomy is that all cite a desire to do justice to the lives of the oppressed as motivation. The same laudable aim can produce different theories, in part, because the label ‘autonomous’ plays a different role for different theorists. Proceduralists often connect autonomy to an array of other goods, such as respect for one’s choices, freedom from coercive intervention, and inclusion in politics. As proceduralists want the politically marginalized to have access to such goods, the capaciousness of their views is a virtue: hence, Friedman argues that a good account of autonomy is one on which “more people can qualify as autonomous” (2003: 23). Others focus on the role the concept of autonomy plays a role in social critique, and its ability to highlight social transformations needed for all to have an equal shot at an autonomous life. Here, the capaciousness of a view is a liability, as it may conceal the full range of ways the social world undermines autonomy, and so be a poor guide to social critique. I understand autonomy in the second sense: a benchmark against which to judge social practices as more or less just. While there are contexts in which concern about coercive intervention is warranted, showing an agent’s autonomy to be diminished does not necessarily license disrespect or exclusion. On this point, I agree with MacKenzie that a failure to be alert to the way the social world threatens personal autonomy can lead to complacency about the ways the world must be changed to ensure justice. While it is important to be aware of how more stringent conditions for autonomy can be used to undermine de jure rights to political autonomy, she writes, these same conditions can explain how relationships or institutions are unjust, as they “impair and restrict agents’ capacities to develop and exercise de facto personal autonomy, even if they possess de jure rights to political autonomy” (MacKenzie 2008: 524).

11 Westlund’s content-neutral, relational view, which I discuss in §3, is an exception here. See Westlund (2009) and (2003).
relationships definitive of autonomy, for instance, this automatically renders a life lived in
deferece to others as non-autonomous, a move that Christman, characteristically, rejects as
“dangerous” and “ultimately problematic” (Christman 2004: 158).

Few deny the causal version of the claim that proceduralists grant. Nevertheless, if
one is primarily concerned with the causal impact of social conditions on an agent’s abilities,
certain threats to autonomy appear more salient than others. The model privileges, in
particular, social forces that are external to agents and have a coercive impact on their
ability to engage in the kind of critical self-reflective process proceduralists take as central
to autonomy. This tacit focus can be seen in the kinds of examples to which proceduralists
turn when reflecting on how a social context can undermine autonomy. Imagining a woman
whose autonomy has been diminished by oppression, for instance, Christman pictures one
who has been “denied all education...systematically punished when expressing curiosity
about alternative conditions” and whose “skills have been narrowly fashioned to accept only
one role” (Christman 2001: 206). Friedman’s discussion similarly reveals an overriding
concern with coercive, external threats: an agent cannot be “overly subjected to coercive
pressures or the controlling power of another person” while engaged in self-reflection and,
once agents have reflectively affirmed their values, the social world cannot severely constrain
their ability to act according to those values (Friedman 2003: 18). While Christman and
Friedman consider different scenarios, the most relevant social forces for both - whether
punishment, another’s controlling interference, or severe constraints on one’s ability to act -
are largely external to agents and possess a coercive dimension.

While the causal model the proceduralist favors can show how a certain class of
oppressive social forces diminishes autonomy, it cannot easily articulate how cultural
meanings and norms do so. For these furnish agents with a symbolic, normative framework
within which they make sense of who they are and what they value, and so appear as more
‘internal’ and less coercive than the kinds of threats on which proceduralists concentrate. It
is also hard to account for how they compromise autonomy without referring to their
Consider Weeks again. Alert only to threats a casual model makes visible, one is hard-pressed to say how her autonomy is compromised: she is not denied an education, subject to coercive control, or facing severely restricted opportunities. And while it is hard to draw firm conclusions about her critical, reflective abilities, it is no glaring deficiency in this sphere that raises the ‘feminist intuition,’ but the way her choice reflects misogynistic ideas that pervade her hyper-sexualized culture. Joel Anderson and Axel Honneth highlight the particular relationship between culture and autonomy when they observe that the “self-interpretive activity central to autonomous reflection” is “fundamentally framed by the semantic and symbolic field in which that reflection occurs” (Anderson & Honneth 2005: 136). They point out, for instance, that it will be hard for a man to value being a stay-at-home-dad if this is, socially, a euphemism for unemployment. Similarly, the case of Weeks primarily raises the question of how her ‘semantic and symbolic field’ has framed what it means to be a woman, powerful, or sexually liberated, and the effect this has had on her self-interpretive activity. Yet this will remain invisible if one is attuned only to causal threats, which take an external, coercive form.

Before turning in §2 to an alternative conception of the social self, I wish to consider the substantivist approach. Many feminist versions of this approach are motivated by a similar concern that attending to a reflective process alone, barring reference to content, conceals the way oppressive socialization compromises autonomy. Hence, substantivists

---

12 Christman and Friedman address this issue, but in ways that do not satisfactorily address the underlying problem. In response to the worry that the oppressed reflectively endorse desires that mirror their oppressive context, Christman adds a historical, counterfactual condition. If an agent either attends to or, counterfactually, were to attend to the socio-historical process by which one of her desires was formed, and she does not or would not resist its development, she is autonomous with respect to the desire. But how are we to interpret this condition? If agents, in fact, reflect on how their desires were formed, this does not address the worry that the internalized status of oppressive norms means they will not feel alienated from them. If the reflection in question is hypothetical, however, it is unclear what grounds the judgment. If we are to imagine how an agent who had never been subject to oppressive socialization would feel, this does seem to entail a self so detached from a social context that it ‘reflects from nowhere.’ See Christman (1991). For another critique of Christman’s attempt to address this worry, see Benson (1991). Friedman (2003: 24-25) responds by pointing out that everyone is subject to limitations of some kind, and so has ‘adaptive preferences.’ Unless a woman’s behavior is ‘so servile’ that she cannot act in accordance with her own deepest values, or becomes subject to coercive intervention, ‘adaptive preferences’ in themselves – including for traditional femininity – cannot rule out autonomy. While I agree our preferences are always shaped by our social world, I think an account of autonomy must be able to draw finer distinctions between oppressive and non-oppressive ways in which this can happen.
deny that autonomy is content-neutral, while invoking ‘content’ in different ways: strong-substantivists directly constrain the content of desires or relationships, while weak-substantivists incorporate content indirectly. Most germane to our purposes are views, such as those Stoljar and Benson develop, that seek to remedy proceduralist shortcomings in capturing the threat of an oppressive culture to women’s autonomy. While Stoljar and Benson put forth strong-substantivist and weak-substantivist views, respectively, both focus on cases in which women meet all proceduralist requirements for autonomy, but have internalized false, oppressive norms, whether those that surround sex and pregnancy (Stoljar 2000: 109) or beauty (Benson 1991). As they have internalized these norms they cannot perceive them as false, and so develop preferences that reflect their oppressive socialization. Stoljar argues that in such cases, a woman’s autonomy is diminished by virtue of the content of the preferences she develops, while Benson argues that it is the way false, internalized norms undermine her ability “to be aware of applicable normative standards, to appreciate those standards, and to bring them competently to bear in one’s evaluations of open courses of action” (Benson 1990: 54), as well as her ability to appreciate the genuine source of her “real strength and value” (Benson 1991: 396). Despite these differences, both agree that the root of women’s diminished autonomy in such cases is that they have internalized oppressive norms they cannot identify as false.¹³

In order to bring out the strengths and limitations of the substantivist approach, I turn to a second case. This is drawn from Susan Bordo’s analysis of anorexia nervosa, an epidemic of self-starvation that emerged among young women in the mid-20th century (Bordo 1993).¹⁴ Bordo situates this phenomenon in a wider cultural context to complicate the assumption that the act of self-starvation is solely a symptom of objective pathology or a desire to emulate thin models. As these interpretations cast women as passive and their activity as lacking deeper significance, they tacitly obscure what Bordo wants to bring to the

¹³ Although Stoljar focuses on the content of women’s preferences, and Benson, their competency in recognizing the source of their value, the difference between these views strikes me as relatively slight. For both, the core problem is that internalized, false norms block women from forming beliefs or preferences that reflect reality; indeed, Stoljar finds much in common between her view and what she terms a strong normative competency view. See the discussion in Stoljar (2000: 107-109).

¹⁴ See also Brumberg (2000) for analysis of historical practices of female self-starvation from the Middle Ages, to the Victorian era and present-day, and Chernin (1981).
fore: the sense in which anorexia represents an “unconscious, inchoate” (Bordo 1993: 175) protest against gendered oppression and a striving for autonomy. Bordo observes that in first-personal narratives, anorexic women betray a visceral fear of inhabiting a mature female body, imagined as an “archetypal image of the female: as hungering, voracious, all-needing, and all-wanting” (Bordo 1993: 160). Their narratives also mirror hegemonic ideas about gender, and the nature of the body and mind. One theme that emerges, for instance, is an association between values of purity, control, transcendence and freedom, on the one hand, and the mind, on the other hand - both coded masculine. Against this backdrop, inhabiting a mature female body appears to necessarily entail dispossessing of valuable, ‘masculine’ traits, such as dignity, self-control, and freedom. And the achievement of these traits appears to depend on the mind’s ability to triumph over the body - coded feminine.

Self-starvation can appear to be irrational when analyzed at an individual level. But Bordo reveals its intelligibility as a way in which young women try to “find honor on the ruinous terms of...[their]...culture” (Bordo 1993: 65). By denying bodily desires, and eradicating signs of female embodiment such as breasts, menstruation and fat, young women aim to embody “an ethics and aesthetic of self-mastery and self-transcendence, expertise, and power over others through the examples of superior will and control” (Bordo 1993: 178). By crafting their bodies, women seek to publicly demonstrate their possession of esteemed ‘masculine’ traits, and to dissociate from the negative traits that attach to female embodiment. In this case, too, women value and pursue autonomy: they want to define themselves in opposition to oppressive, alienating ideas about who they are. Yet the way in which they do so is self-undermining and, in certain respects, reinforces patriarchal ideas about women. In “disturbingly concrete, hyperbolic terms” (Bordo 1993: 169), anorexia colludes with traditional gender norms: women deny their own desires, remain physically fragile, and take up little room in the public sphere. And although young women may feel “deeply attracted to the aura of freedom and independence suggested by the boyish body ideal” (Bordo 1993: 160), their pursuit of this ideal leads to serious physiological problems, and the expenditure of time, energy and attention on a narrow aim that impedes the development of their individual talents. Paradoxically, the act of self-starvation can be seen
to re-inscribe the very normative ideal of femininity that, on another level, it aims to counter.

A substantivist approach is well-equipped to articulate what goes wrong in this case: women women perceive and appraise themselves in light of oppressive gender norms such that, even with reflective abilities intact, they ends they formulate bear the imprint of such norms. Yet the substantivist approach also risks obscuring the ambiguity that cases of thwarted autonomy exhibit - and so what ‘goes right.’ For inquiry into the content of the women’s desires and values yields an ambiguous result: not only do they latch onto and pursue the right value of autonomy, but they also display a deep ambivalence about the oppressive meanings and norms that imbue their culture. It seems, then, one-sided to say that they have internalized them and so cannot perceive them as false. If the women were entirely uncritical of the idea that freedom and independence are male traits, for instance, it would be hard to make sense of how their activity could also be a nascent rebellion against norms that cast women as unable to achieve such traits. Second, if one imagines that the women have internalized oppressive norms, one risks inadvertently reinforcing the very image Bordo wants to avoid, of women as merely passive repositories of the oppressive meanings and norms that circulate in their culture. Indeed, on the basis of such an image, it is hard to see how women could gain critical purchase on oppressive norms or how, in attempting to fashion an ‘honorable’ self out of a ‘ruinous’ culture, they could recognize and pursue the value of autonomy itself.

Capturing the case’s ambiguity and critical edge requires one to reject the assumption that women conform to patriarchal practices because they have internalized, and so uncritically accept sexist norms. To bring this into view, one must attend to women's practical, embodied expressions of opposition to oppressive meanings and norms that, in fact, characterize their social world. For while it is tempting to see the women Bordo discusses as unable to appreciate normative standards in the right way, the culture in which they live really does connect masculinity to the mind and vaunted traits, and femininity, to the body and degraded traits. Rather than see the women as normatively incompetent, one can see their activity as reflective of an acute awareness of prevailing normative standards in
their social environment and, on that basis, reasons they have for acting as they do.\textsuperscript{15} Uma Narayan argues, along similar lines, that women’s conformity to sexist practices need not reflect their acceptance of those practices, but their attempts to pursue what they want in light of realistic judgments they make about their options. (Narayan 2002: 425).\textsuperscript{16} This activity of ‘bargaining with patriarchy’ lies in between a “dupes of patriarchy” (Narayan 2002: 425) model, on which women uncritically accept oppressive norms, and a “prisoners of patriarchy” (Narayan 2002: 419) model, on which women are coerced into compliance. Similarly, women’s outward conformity to patriarchal norms in cases of thwarted autonomy need not indicate a deficiency in their normative competency. Rather, one can interpret these cases as ones in which women try to get what they want - autonomy - in light of accurate judgments they make about the nature of their social world, and the normative standards at play within it.\textsuperscript{17}

\textsuperscript{15} Similarities exist between Benson’s normative competency view and an expressivist view. For Benson, free agency depends on the extent to which agents are aware of the social norms that govern their self-disclosure (Benson, 1990). An expressivist view also emphasizes the idea that dominant social meanings and norms mediate the way agents express their sense of self. Although Benson points out the way oppressive social norms place women in double-binds, his view risks conservatism, as it ties free agency to the ability to act in ways that cohere with social norms; Benson takes this to point to the need for consciousness raising. An expressivist view can avoid this risk because it focuses not just on women’s awareness of prevalent social norms, but what they seek to express about themselves and the process through which they do so. Although the women in cases of thwarted autonomy track, and act in light of, ‘socially real’ aspects of the world, such as the symbolism that attaches to female embodiment or the association between masculinity and the mind, an expressivist view also considers what they are trying to achieve through their activity and the extent to which, as individuals, they reflect on and refine their activity over time.

\textsuperscript{16} See Khader (2012) for another account that makes ‘bargaining with patriarchy’ central.

\textsuperscript{17} Cases of thwarted autonomy are, similarly, not well understood in terms of external coercion, or internalized norms. My interpretation of the cases shares much in common with Narayan’s rejection of the ‘prisoners’ and ‘dupes of patriarchy’ dichotomy. Indeed, as I elaborate in §3, the women’s attempt to express a desire for autonomy within the strictures of their social environment can be understood as a version of ‘bargaining with patriarchy,’ albeit transposed onto the hermeneutic sphere. One note of caution is in order, however. ‘Bargaining’ can incline one to think in instrumental terms, such that a women’s aims or interests may be understood as distinct from, and prior to, to the practices through which they fulfill them. A woman who veils in order to secure a better marriage, while not invested in the practice itself, is doing “the best she can to advance her interests,” (Khader, 2012: 305) where her interests can be identified apart from the practice itself, which is a means to an end. Cases of thwarted autonomy are not instrumental in this sense, as women develop a sense of their desires through the practices in which they engage, and insofar as the specific symbolism of the practice matters. Narayan’s discussion of the non-instrumental relationship women have to practices that provide them with social, religious, and communal identities is closer to the dynamic I have in mind. Yet this also leaves out the way in which women in cases of thwarted autonomy are trying to actualize a not-yet settled sense of who they are. The ‘bargaining with patriarchy’ that occurs in these cases may be best understood as occupying a middle ground between instrumental and non-instrumental ways of relating to practices.
This section's aim has been twofold. First, I introduced a novel set of cases, in which women pursue autonomy, but in ways that reinforce patriarchal norms. Second, I drew out underlying socio-ontological assumptions of two dominant approaches and considered the extent to which they capture the cases. The upshot is that a positive view must accommodate: (1) the way an oppressive culture furnishes agents with a symbolic, normative frame through which they engage in self-interpretive activity and (2) the sense in which agents can engage in meaningful, practical opposition to oppressive norms, even if not evident from outward behavior. An expressivist conception of the social self, I shall now show, meets both desiderata.

II. Expressive Activity

Autonomy involves self-understanding: the ability to identify with or feel alienated from desires, ends, and values requires some sense, even if not entirely transparent, of who one is. An expressivist view, which intervenes at the level of social ontology to inquire into the necessary conditions for self-understanding, argues that it is only by expressing desires and values in a social medium that we come to discover who we are. This view enables one to bring the wider sphere of ‘expressive activity’ into view when assessing an agent’s autonomy. In this section, I briefly sketch this idea before applying the framework to the cases of thwarted autonomy in §3.

The expressivist view begins from a broadly Hegelian insight into the nature of self-understanding: that agents are not immediately transparent to themselves, but must engage in a certain kind of social activity to achieve this state.18 As Charles Taylor glosses this idea, self-understanding is not immediately given but an achievement: it is, he writes, the “fruit of an activity of formulating how things are with us, what we desire, think, feel, and so on” (Taylor 1985: 85). Our ideas, emotions, and desires are initially inchoate and elude our full grasp. Reflection alone, however, does not yield clarity. To know, and be able to identify with or feel alienated from, desire or values we must publicly articulate our vague sense of what

---

18 The account of expressive activity that follows draws inspiration from Hegel’s conception of action and its connection to agency. I have benefited, in particular, from Taylor (1985), Pippin (2010) and Anderson (2010).
these are in the social world. A paradigmatic example of this expressive activity is that of writing an essay or a journal entry. One often begins with a hazy sense of what one thinks or feels. As the idea of ‘writing to find out what you think’ captures, however, it is by engaging in the activity itself - with its demand that one find the right words - that one develops a more lucid and concrete sense of one’s thoughts and emotions. It is through this activity of articulation that one arrives at self-understanding.

Expressive activity can aid self-understanding in a passive sense, as when one’s action contradicts one’s self-image. I may fervently avow commitment to a political cause but, if I never attend a meeting and spend all my time playing music instead, I may conclude that, contrary to what I wish to believe about myself, I in fact value music more than politics. I can also, however, actively strive to articulate my desires and values in the world, and gradually refine this articulation as my sense of myself emerges over time. As Taylor notes, at first expressive activity is “relatively unreflecting” (Taylor 1985: 91). But through the process of reflecting on and realizing deficiencies in the form my activity takes, I may seek out more refined modes of expression. I might initially express the value I place on justice by posting on social media, for instance, but come to feel that this mode of expression is too superficial, detached from the world, or egocentric. Realizing that its form is not adequate to what I want to express, I may instead join an in-person activist group, as a better way in which to embody this value. This activity, of expressing, reflecting on, and revising the public articulation of our desires and values, yields self-understanding.

What matters for our purposes is the image of the social self that corresponds to this picture. First, as expressive activity requires a social medium, the social world will partially shape the form one's outward expression takes. The way one expresses anger in a private journal, for instance, will differ from how one does so on an online platform set up to incentivize attention-grabbing and polemical forms of expression. More broadly, the intelligibility of my activity, and the form it takes, will depend on social meanings, norms, and others’ interpretations. In order for an act of rebellion to be intelligible as such, for instance, it must be visibly at odds with status quo values, behaviors, and modes of appearance that are themselves culturally and historically specific. Silicon Valley's casual
attire can only be read as reflective of an ethos that valorizes ‘disruption’ and rejects ‘business-as-usual,’ not just an arbitrary fashion choice, against the backdrop of a status quo characterized by staid suits; it could not have this same meaning in 15th-century France. The way others interpret my expressive activity also matters to what it can signify. If what I take to be an act of rebellion is widely interpreted as a shallow attempt to stand out, for instance, there is a sense in which I have failed to express what I wanted to; this in turn may lead me to reexamine and refine my activity. If this points to the way in which the social world constructs expressive activity, highlighting the sense in which individuals are not “self-sufficient bearers” (Anderson 2010: 18) of social meaning, it does so only partially. For agents also play an active role in expressing their desires and values in a way that is not merely determined by the social world. The particular desires and values I express, and how I choose to do so will reflect my individuality; as Anderson observes, an agent’s “character, history, mood, energy, actions, and reading of her predicament play a profound part in influencing what she values, especially in influencing which practices, roles, and relationships she will make her own” (Anderson 2010: 25).

The outward form of one’s expressive activity is partially constituted by one’s social context. But insofar as this activity furnishes us with self-understanding, it makes little sense to imagine that one can achieve such self-understanding by standing wholly apart from one’s social context, or that this context has no effect on who one becomes. In particular, the practices in which one engages to express some desire or value provide a ‘semantic and symbolic’ frame with which to interpret one’s desires, emotions, relationships, and so forth. Through this activity, then, what are initially inchoate desires and values come to be imbued with socially mediated content. To express my rejection of mainstream values in a way intelligible in terms of my social context, for instance, I may gravitate towards a subculture whose practices represent the most visible rejection of such values. Engaging in the subculture’s practices, however, will leave neither my initial desire nor my conception of myself unchanged. Perhaps I come to strongly identify as a member of the subculture, and cannot imagine who I would be apart from it, or view outsiders to the group as shallow or morally corrupt. And my initially inchoate desire to reject mainstream values may acquire a
specific content: to wear all-black, for instance, follow a vegan diet, or become an adherent of an extreme political ideology.

With this brief sketch of the expressivist view of the social self, we can see one way of conceptualizing the relation between agents and the social world. Although agents are not determined by their social context, the expressive activity through which they arrive at self-understanding is socially mediated. Socially available meanings and norms therefore shape not only the outward form their expressive activity takes, but also their conception of themselves and what they desire. I turn now to show how this expressivist view of the social self can illuminate the cases of thwarted autonomy from §1, and what it reveals about a social conception of autonomy.

III. Toward an Expressivist View of Women's Autonomy

This view invites inquiry into the wider sphere of expressive activity. What is an agent trying to express? In what ways do cultural meanings and norms shape the form her expression takes? How does her expressive activity structure the relationship she has to herself? In this final section, I consider what answering these questions, in cases of thwarted autonomy, can reveal about what it means to conceptualize autonomy in a social register. In the cases from §1, women do express a desire for autonomy, as their activity reflects a desire for self-definition and direction over their own lives.19 But initially this desire is inchoate. It may manifest only as a pressing discomfort with how others perceive them, or frustration at the way this restricts their capacity for self-definition. Reaching - at first relatively unreflectively - for a way to express this as-yet unformed desire in the world, the women encounter a set of meanings and norms that structure its intelligible expression. Following Marilyn Frye’s (1983) definition of oppression, these meanings and norms are oppressive:

19 Autonomy can refer to the expressive activity in which agents engage: if expressing the value I place on artistic practice enables me to understand and develop my commitment to art, this can be understood as a development of my autonomy. I refer, however, to the women as ‘expressing a desire for autonomy’ or ‘expressing the value they place on autonomy.’ With this location, I intend to draw attention to what is important about these cases: that the women do not express any old desire, but specifically, a desire to direct their own lives.
they set up double-binds that penalize women for any choice they make. The result is that women face particular challenges in trying to express a desire for autonomy.

Artistic, cultural, and everyday practices often portray women as sexual objects who exist for the pleasure of others, and as dispossessed of traits like independence, powerfulness, and self-assertion. When women do exhibit such traits, they often face explicit or implicit forms of social censure. Or such traits are conflated with sexual objectification, as when female celebrities claim to be empowered by virtue of their conformity to highly sexualized roles. This puts women in a double-bind. If they accept the status of sexual object, they can avoid social censure, but they then become complicit in their own self-effacement. But if they reject this complicity in favor of a more robust autonomy, they expose themselves to stigma, critical scrutiny or, at best, a tenuous recognition. Part of what perpetuates this double-bind is the freighting of female embodiment with social meanings of being passive, object-like, or ungovernable that put it in tension with autonomy. The prevalence of such meanings does not determine the relation a woman will have to her own body, or wholly block her expression of autonomy. But it does make it challenging for women to escape or re-signify such meanings, or to straightforwardly express a desire for autonomy.

Cases of thwarted autonomy can be understood as attempts to express a desire for autonomy in a social context that problematizes this act of expression for women. As this context casts the female body as at odds with autonomy, it is unsurprising that the women try to express their desire for autonomy by cultivating a relationship to their own bodies that can intelligibly signify opposition to oppressive meanings that would otherwise be simply imposed on them. Rather than mount a collective challenge to such meanings by rejecting broader assumptions about what it means to exist in a female body, women try as individuals to express a desire for autonomy within this symbolic and normative framework. By drastically shrinking her body, for instance, the anorexic woman defines herself in opposition to negative meanings that would otherwise attach to it. If the female body is an

---

20 For elaborations of this mainstay of feminist thought see De Beauvoir (2011) & Wolf (1990).
obstacle to autonomy, she rejects and reshapes her body to publicly signal her achievement of this value. Weeks does not reject her body so much as try to actively appropriate meanings that attach to female embodiment. By actively positioning herself in relation to sexually objectifying meanings, she can feel she has chosen them for herself—bearing out Simone de Beauvoir’s perceptive analysis of the ‘woman in love’ who tries to “overcome her situation as inessential object by radically assuming it” (De Beauvoir 2011: 684).

On the one hand, by seeing these cases through an expressivist lens we can heed Khader’s recommendation to look for “signs of agency in oppressed people—even when that agency is not readily visible” (Khader, 2012: 313). By foregrounding the questions of what the women seek to express, and how cultural meanings and norms shape this expressive activity’s outward form, the framework allows us to reinterpret women’s apparent acquiescence to gendered subordination. Women need not be ‘dues of patriarchy,’ in thrall to images of beauty or blindly acquiescent to their own sexual objectification. Rather, they can appear as trying to find a way to express their desire for autonomy in a world that gives them few good options for doing so—a version of ‘bargaining with patriarchy,’ transposed onto the hermeneutic sphere. As the expressivist view takes the sphere of expressive activity as central, it also permits consideration of how the women engage in such activity and, in doing so, casts autonomy as graded—not something one can possess more or less of, but an activity one can do more or less well. Shying away from expressive activity altogether can lead one to develop only an attenuated sense of self, while doing so in an overly confident, unreflective manner may mean one fails to reap the rewards of self-understanding such activity can provide. The women in the cases from §1 are actively attempting to express a desire for autonomy in the world. As such, we can recognize that they are engaged in the kind of activity that could, in principle, form the basis of an autonomous life.

On the other hand, and alongside these positive ‘signs of agency,’ the expressivist view can articulate the sense in which oppression thwarts women’s autonomy in these cases. By making the socially mediated character of expressive activity and the self central, it can reveal how an oppressive social world will inhibit or distort attempts to express autonomy. The remainder of the paper will spell out how it does so by pointing towards the ways in
which an expressivist view both borrows, and departs, from a class of relational views
developed by Andrea Westlund (2003, 2009) and Catriona MacKenzie (2008), among others. The distinctive claim of these views is that one’s relationship to oneself, and one’s ability to be autonomous, is constituted in and through intersubjective relationships to others - an intersubjective realm that, as Westlund notes, does not “map onto any straightforward distinction between what is internal and what is external to the agent” (Westlund 2009: 33). Similarly, key to the expressivist view is the claim that social meanings and norms structure the outward form expressive activity takes, and the self-understanding one develops through such activity. Here too, the social world mediates and constructs the relationship one has to oneself, the site of autonomy. While the following discussion will be necessarily truncated, it will be useful, then, to consider what an expressivist view can take up from these views, and how it modifies them.

For Westlund, autonomy is a critical, self-reflexive stance towards one’s own commitments. It is constituted through dialogue with others and depends, in particular, on the extent to which one is ‘answerable’ for one’s commitments in the face of others’ critical challenges (Westlund, 2009: 33). Westlund’s proposal sheds light on what expressive activity must entail if it is to lead to autonomy. In trying to make sense of my own activity, which may be incoherent or self-defeating at first, I look to the interpretations and challenges (Taylor 1985: 89) of others. The way I engage with such reactions matters. Others may, for instance, tell me that the way I express the value of justice is alienating or merely reflects my narcissism. I may ultimately reject these interpretations. Nevertheless, being receptive to, and willing to engage with, such challenges both indicates that I have the right kind of self-reflexive relationship to my own commitments, and affects the extent to which I can revise my activity to better reflect what I am trying to express.

Yet solely focusing on the interpersonal realm can obscure more systemic forms of oppression that do not take a dyadic, relational form. Hence, the scope of the social world must be expanded beyond the dialogical realm to include dimensions not easily

---

21 For a moving discussion of relational autonomy in the context of trauma, see Brison (2002).
characterized in relational terms. If the social world lacks certain hermeneutical resources, for instance, attending only to one’s willingness to engage in justificatory dialogue will be insufficient. When homosexuality was, on dominant interpretations, seen as a psychological disorder in need of treatment, for example, this would have appeared as an analyst’s critical challenge to a patient, with a narrative about their activity should be revised - namely, towards heterosexuality. One can describe this interaction as having an ideological dimension, as the analyst’s interpretation serves to naturalize unjust social arrangements.

and receptiveness to critical challenges of others may mystify the nature of an agent’s activity to herself, and lead her to attempt to revise it in ways not genuinely emancipatory.\textsuperscript{22}

Women’s reactions to misogyny are, along similar lines, often construed as individual, psychological defects, and as disconnected from socio-political conditions. A woman who receives the message that her activity is rooted in an individual flaw may ‘work on herself,’ but if her activity is, in fact, rooted in oppressive social conditions, she will encounter difficulties in fully understanding her own activity or revising in liberating ways. The point is not that engaging in ideologically laden forms of justificatory dialogue necessarily diminishes an agent’s autonomy; it is that the wider social forces that structure the dialogical realm must also be brought under consideration in reflecting on how oppression undermines autonomy.

For MacKenzie, autonomy is a normative authority over one’s life that is grounded not just in answerability to others, but in attitudes of self-esteem, self-respect, and self-trust, and in processes of self-interpretation. Insofar as it is within intersubjective relationships of recognition that we figure out who we are, what matters to us and, in Honneth’s words, can see ourselves “as beings whose needs, beliefs, and abilities are worth being realized” (Honneth, 2012: 41), autonomy is sustained through relationships of recognition, and undermined by disrespect or violations of one’s physical integrity.\textsuperscript{23} Again, this insight captures a key intuition about why the cases in §1 are ones in which autonomy is thwarted. Anorexia and misogynistic pornography both raise the ‘feminist intuition’ not only by virtue

\textsuperscript{22} See Haslanger (2012) and (2017) for discussions of ideology.

\textsuperscript{23} Westlund’s view shares much in common with Honneth’s view of autonomy as sustained by intersubjective relationships of recognition. See Honneth (2012).
of the way they mirror patriarchal ideas about women, but because both also entail forms of self-abnegation, physical harm, and, in Weeks’s case, degradation by others. Expressing a desire for autonomy through these particular practices may erode the kind of affective attitudes agents need to see themselves as “beings whose needs, beliefs, and abilities are worth being realized” (Honneth 2012: 41). MacKenzie does connect relationships of recognition to social practices and institutions. Nevertheless, the concept of recognition can still incline our thought towards the interpersonal, and so I wish to conclude by pointing to two reasons the normative structure of practices themselves must be equally at the fore.24

A brief, stipulative definition: a practice is an activity governed by constitutive rules that must be followed for the activity is to be performed correctly, and out of which emerge social meanings, norms, roles, and relationships. Emphasizing the structure of practices can, first, capture cases in which direct failures of intersubjective recognition are not apparent. The forms of self-abnegation and physical harm anorexia involves, for instance, do not inhere in intersubjective relationships; this is, rather, an activity individuals carry out alone. The features MacKenzie takes as indicative of diminished autonomy show up when one views this phenomenon as a practice: one that involves rules one must follow to perform it ‘correctly,’ which give rise to certain meanings and norms. One of Bordo’s observations is that by crafting their bodies in a certain way, young women strive for ethical and spiritual qualities, not just aesthetic ones. The point is borne out by online forums dedicated to ‘pro-anorexia’ discourse, a central theme of which is the pursuit of ‘perfection’ - a state at once physical and imbued with quasi-spiritual properties - and the worthlessness of those who fail to achieve it. As this is a pursuit whose goal is never reached, participating in it involves viewing oneself in a lowly light, forever falling short of an ideal. Combined with the denial of basic bodily needs the practice involves, over time, participation may erode the attitudes of self-trust, self-esteem, and self-respect one needs in order to be autonomous - an erosion that does not necessarily appear if we focus on failures of intersubjective recognition, but

24 MacKenzie points out, for instance, that the authorization to speak for ourselves are sustained by practices, and argues that the “basic social, legal, political, and economic institutions” in a just society furnish the “recognition basis for its citizens to realize their autonomy.” See MacKenzie, 2008: 524.
that does if we look to the structure of the practice itself. If one concentrates on failures of recognition in the interpersonal sphere, therefore, it may not show up

Second, foregrounding the structure of practices can orient social critique in the right way. MacKenzie argues that social institutions meet the demands of justice when they involve the relationships of mutual recognition that sustain autonomy. The rules of practices can also license failures of recognition, however, such that these must be brought into view as a root cause of the way oppression undermines autonomy. The case of Weeks does involve a failure of recognition in the interpersonal sphere, for instance. But social critique will be insufficient if it solely targets relationships of disrespect without looking, equally, at the way social practices enable them. Feminists have argued that the constitutive rules of pornography as a practice eroticize a submissive feminine role and dominant masculine role, preventing women’s refusal of sex from having the right force (MacKinnon 1987; Langton 1993). The normative structure of the practice establishes a particular dynamic, in which participating as the female partner involves adopting a sexualized, submissive role and being dominated by others: to do otherwise would be to perform the practice ‘wrong,’ or be engaged in a different practice altogether. If, in short, it is the structure of a practice that licenses failures of recognition or the transgression of an agent’s physical boundaries, this structure itself must be an object of social critique.

Conclusion

Feminist philosophers agree that autonomy must be understood in a social register if the concept is to be useful for the diagnosis and redress of patriarchal oppression. Less consensus exists about the way the social world must figure in autonomy. In this paper, I have introduced the under-theorized phenomenon of thwarted autonomy to clarify the socio-ontological issues at stake in this debate and to motivate an expressivist view of the social self, as a foundation for an account of autonomy. This view emphasizes that we develop self-understanding by practically embodying our inchoate desires and values in the social world, and so the way in which the social world shapes the form one’s expression takes, and one’s self-understanding. On this view of the social self, we can reinterpret
women’s outward conformity to gendered subordination as an attempt to actualize autonomy, in a social context that provides women deficient vehicles through which to do so. I have also sketched some ways in which oppression undermines women’s autonomy, on this view. In the absence of the right hermeneutical resources and social practices, women may be unable to understand or revise their activity in ways that are genuinely liberating. They may participate in practices that encourage them to see themselves in a submissive or abject light, and so erode the affective attitudes they need to exercise autonomy. While a full account of an expressivist view of autonomy must await future elaboration, I hope to have brought into view a wider sphere of what matters in making sense of oneself.

Acknowledgments

For helpful and generous comments on previous versions of this paper, I would like to extend my thanks to César Cabezas, Jonathan Fine, Eric Bayruns García, Yarran Hominh, Megan Hyska, Hannah Martens, Annette Martín, Philip Yvaure, Daniel Brinkerhoff Young, and the editors and referees at Ergo. I am also grateful to audiences at Northwestern University, Binghamton University, the University of Edinburgh, the Social Philosophy Workshop in Chicago, and the Social Ontology Conference in Vienna.

References


