

How can consciousness be false?

Alienation, simulation, and mental ownership

Matteo Bianchin

Philosophy and Social Criticism (forthcoming)

penultimate draft

1. Understanding Alienation

Alienation has been recently revived as a key concept in critical theory by merging classical Hegelian and Marxian views with existential elements that draw on Heidegger and, partially, on Sartre (Jaeggi 2014, 12 ff., 82 ff.).¹ What enables merging otherwise divergent understandings is the idea that alienation involves a peculiar experience of self-estrangement: a sense of not living one's own life, the inability to identify with one's own actions, a failure to recognize one's own existence. In this view, the mark of alienation is a distinctive failure in relating to the relationships we entertain with ourselves as well as with the natural and the social world, which results in a pervasive sense of meaninglessness, powerlessness, and indifference (Jaeggi 2014, 25; Rosa 2019, 186).² Alienation is not a failure in relating with the subjective, the objective, and the social world, but rather a failure to relate to being related. It is not that

¹ Honneth (2012, 82 ff., 103 ff.) takes a similar line on reification.

² While both Jaeggi and Rosa depart from Marx in several ways, this abstract characterization is consistent with Marx's views (see Elster 1986, 49 ff.; Wood 1981, 7 ff.)

such relationships are absent in alienated subjects; rather, they are lost to them, as alienated subjects are unable to experience them as their own. Alienation thus seems to entail a peculiar disruption of the self that occurs as we fail to appropriate the relationships we do entertain with ourselves, with others, and with the world, which results in experiencing them as “external, unrelated, non-responsive” — in a word, “mute” (Jaeggi 2014, 37 ff.; Rosa 2019, 187). This involves a disturbance in our desires, beliefs, emotions, and actions that does not bear on their content or attitudinal features, but on the sense of ownership that goes along with them. Being alienated involves experiencing ourselves as estranged from the desires, the beliefs, the emotions, and the actions we go through — failing to experience them as our own:

[...] self-alienation will be conceived [...] not as falling away from one’s own essence but as a disturbed relation to self, as a disturbed relation to our actions, desires, projects, beliefs [...] when alienated, we are alienated from something that is *simultaneously our own and alien* (Jaeggi 2014, 47-48).

In this paper, I contend that alienation involves a disruption in the sense of mental ownership that comes with the first personal, pre-reflective self-awareness of being a subject of experiences, attitudes, and actions, and I argue that social factors play a structuring causal role in the process that brings about alienation. In this section, I introduce the theme and explain why it is important to focus on the mechanisms that underlie alienation. In the second section, I maintain that understanding how alienation works is crucial to make sense of false consciousness. In the third section, I consider the relevance of mental ownership to explaining alienation, and I discuss existing evidence about whether and how it can fail. In the fourth section, I argue that disturbances in the simulation routines that support social cognition might

underpin alienation, and I outline how social factors might play a structuring causal role in this connection.

It is common ground that alienation seems to be a paradoxical condition. We are only alienated as *we* become alienated *from ourselves*. But how can sense be made of the idea that one can be alien to oneself, that one can be *not* oneself, that our life can be “not our own”? (Jaeggi 2014, 44). The paradox here differs from the well-known paradoxes involved in self-deception, as it does not concern the contents and the dynamics of propositional attitudes. In particular, it does not result from believing and disbelieving, or desiring and repelling the same thing (Davidson 1985; Galeotti 2018). Talks of alienation rather suggest that we can happen to experience the attitudes, emotions, and actions we undergo as if they are not, or not fully, our own. Alienation diverges from self-deception because it does not concern how we can deceive ourselves about *what* we believe, desire, feel, or do but how the beliefs and desires we entertain, the emotions we undergo, or the actions we carry out may fail to be our own.

Consider Jaeggi’s case for the lack of agency. Jaeggi’s example concerns a young academic whose “slightly chaotic life” eventually settles and acquires “all attributes of a completely normal suburban existence”. Although no coercion or heteronomy is involved in the process, the life he leads unperceptively turns to look “not to be his own”, as if “an alien power was [...] working through him” (Jaeggi 2014, 52-53). On the one hand, life seems to have taken “a dynamic of its own”, as the agent “is not present in his life”; on the other hand, social relations “rigidify” so that he becomes “a passive and no longer an active participant in the relations in which one lives” (Jaeggi 2014, 54, 56-57, 59). According to Jaeggi, alienation here involves a failure to understand oneself as the subject of one’s own actions (Jaeggi 2014, 51). On the subjective side, it affects agency by affecting the subject’s access to itself. Alienated subjects fail to relate to their own actions because their “accessibility to the self” is disturbed, where agency requires the agent to be “accessible to oneself in what one experiences and does” (Jaeggi 2014, 66). On the objective side, a variety of social causal factors are supposed to

alienate subjects from their own actions: the anonymous command of social conventions and late capitalism's "pathologies of work" recur in this connection as paradigmatic examples (Jaeggi 2014, 67; Jaeggi 2017, 66).

Jaeggi consequently conceives alienation as a "relation of relationlessness" (Jaeggi 2014, 1) to capture a structural feature of self-estrangement that underlie both Marxist and existentialist conceptions: a failure to relate to the relations we *do* entertain with the subjective, the objective, and the social world that turn them "*simultaneously our own and alien*" (Jaeggi 2014, 49). On this reading, self-estrangement involves a peculiar disruption in the sense of self which is not in itself specific to alienation as a social phenomenon. Jaeggi herself occasionally compares the latter with the psychological description of a clinical condition:

A person who is alienated from herself has (as a psychological description of a clinical symptom would have it) lost a relation to her own feelings, desires, and experiences and can no longer—even to the point of spatiotemporal disorientation—integrate them into the way she experiences her own life. She is alien to herself in what she wills and does. Incapable of experiencing herself as an actively structuring force, she feels unable to have any influence on what happens to her, which instead she experiences as something alien. (Jaeggi 2014, 44)

Disturbances in one's relatedness to one's self and the world are well documented in the clinical literature and commonly traced back to a disruption in the basic sense of self that goes along with being a subject of experience, action, and thought – a "minimal self" conveyed by pre-reflective self-awareness (Ciaunica, Charlton, Farmer 2020; Sierra 2009, 24 ff.; Billon 2017; Gerrans 2019; Damasio 2000; Zahavi 2005, 106, 124 ff.; Kriegel, Zahavi 2016). Commonly associated with severe psychopathologies, such disturbances are diagnosed as depersonalization disorders when chronic and debilitating. Phenomenologically, they involve

absent or deficient body-ownership feelings, agency deprivation, emotional numbing, anomalous subjective recall and imagery, and a general sense of detachment from the world. Explanatorily, they are taken to result from a disturbance in the “minimal self” that affects our ability to relate to our own mental and bodily processes as well as to our cognitive, emotional, and agential relations with the world and with others.

Clinical literature focuses on acute manifestations, but milder cases are occasionally elicited by drug or alcohol intoxication, sensory or sleep deprivation, and extreme physical illness or fatigue (Sierra 2009, 44 ff.). Moreover, surveys document a significant prevalence of transient episodes of depersonalization in the general population and consonant first-personal reports of anomalous self- and world experiences have been recently collected in a qualitative study whose participants were screened for the absence of mental health comorbidities, yet scored higher than 50 on the Cambridge Depersonalization Scale (Ciaunica, Pienkos, Nakul, Madeira, Farmer, H. 2022). Participants in this study typically report not being “present” to themselves, lacking ownership over experience and thought, empathy deprivation, emotional numbing, and blurred boundaries between self and others (Ciaunica, Pienkos, Nakul, Madeira, Farmer, H. 20228 ff.).

Interestingly, depersonalization in the general population does not categorically differ from clinical depersonalization, but is transient, milder, less frequent, and includes a narrower range of symptoms. What marks off clinical cases is a matter of degree. As one moves from one extreme of the continuum to the other, episodes are more intense, frequent, lasting, and encompass a wider range of symptoms. The difference between clinical and “normal” cases is a difference in severity that depends on the structural conditions under which depersonalization is induced. While clinical depersonalization is usually a comorbid manifestation of other psychopathological conditions, non-clinical depersonalization is elicited by contingent causal factors like sensory and sleep deprivation, drug or alcohol intoxication, extreme physical illness or fatigue (Sierra 2009, 47-48).

In this paper, I will argue that this difference can be cashed out in terms of a well-known distinction between structuring and triggering causes, and that the same distinction accounts for the difference between alienation and clinical depersonalization. What marks off alienation as a distinctively social phenomenon is that the structuring causes of the relevant disruption in the sense of self are individuated by exogenous social conditions rather than by psychopathological conditions internal to the subject. Therefore, its phenomenal manifestations can be sensibly expected to be milder and narrower than in clinical depersonalization, but more frequent, intense, and lasting than in non-clinical cases elicited by merely contingent causal factors. Schematically, there are thus three levels at which alienation can be analyzed, namely the phenomenological, the mechanistic, and the (social) structural level, where the latter marks off alienation as a distinctively social phenomenon.

Phenomenological level. The phenomenology of alienation does not qualitatively differ from the phenomenology of clinical depersonalization, although it likely differs in severity.

Mechanistic level. The mechanism that underlies the phenomenology of self-estrangement is common to alienation and clinical depersonalization: both involves a disruption in the sense of self, although severity can be expected to differ.

Structural level: the disruption in the sense of self is underpinned by exogenous, social structuring causes in alienation, by endogenous psychopathological structuring causes in clinical depersonalization – which explains the variation in severity.

Section three and four account for alienation along this framework and suggest that it enables telling apart consistently the subjective and the objective dimensions of alienation. Sorting out the phenomenological, the mechanistic, and the structural features of alienation

locates their causal relations so as to match both current readings and the original Marxian view that workers' self-estrangement mirrors the structural domination of an alien power that "rule him by his own actions": workers fail to perceive their labor activities as their own because the subsumption of living labor under capital turn them into a mere "link of the system" and convert social relations into relations between things (Marx 1993: 156-157, 453, 693-94; Elster 1986a, 76-77, 103 ff., 265, Cohen 1978, 120).

In the concluding section, I draw two implications that are relevant to social and political philosophy. First, on this reading alienation is subject to a social structural explanation rather than to a purely psychological account. Second, overcoming alienation consequently involves social critique and transformation rather than psychological diagnosis and therapy. The upshot is that alienation critique cannot be divorced from a social structural explanation that traces it back to structural domination or injustice.

A main trouble for a theory of alienation is to dissolve the appearance of a paradox by accounting for how we can fall away from ourselves, experience ourselves as alien to us, and keep ourselves unrecognized across our own life and action in such a way that the beliefs, desires, emotions, and actions we go through look like they are not our own. In what follows, I advance an account of alienation that bears on the mechanisms that bring it about. This is not meant to contribute a definition or a phenomenology of alienation, which I take to be settled along the lines sketched above, but rather to understand how it can be explained. Thus, I will not take a stand on how non-alienation should be normatively understood, be it in terms of autonomy (Forst 2017), narrative intelligibility (Jaeggi 2014, 170, 179), authenticity (Rosa 2013, 317-318), or resonance (Rosa 2019, 186). Answering the normative question is admittedly crucial to a theory of alienation, yet it provides virtually no information about how alienation works, which seems fundamental to dissolve the paradox and establish alienation as a social and psychological reality, to single it out from behaviorally close phenomena, and possibly to measure the range of its effects. I only retain Jaeggi's and Rosa's negative

characterization, according to which alienation involves a failure to appropriate one's own relations with the self, others, and the world, as it provides a minimal common factor that applies across diverse normative understandings of non-alienation.

The rationale to focus on how alienation works is twofold. First, it is far from obvious that talks of alienation pick out any social or psychological fact. The philosophical vocabulary of alienation might merely voice a declining modern mythology of the self that epicyclically revolves around worn-out ideas of integration, autonomy, and authenticity.³ Moreover, the meaninglessness, powerlessness, and indifference routinely associated with alienation might be read as common by-products of ordinary mental disorders rather than as expressions of a distinctive condition that calls for a *sui generis* theoretical classification. Therefore, a major demand placed on a theory of alienation is *how* the paradox of alienation can possibly track a specific kind of social or psychological facts. Locating the mechanisms at work addresses this demand. More specifically, I suggest that alienation can be read as involving a mild form of depersonalization that can be disentangled from clinical cases by discerning the relevant etiological differences.

Second, it seems desirable that detecting alienation as a social and psychological fact does not depend on prior answers to controversial normative questions. Locating the mechanisms and the social causal factors at work provides the tools for detecting alienation that is free from prior commitments to substantial normative views of non-alienation. As we will see, alienation causally depends on the obtaining of specific social conditions that factor as structuring causes in the process that brings it about and are likely to involve structural injustice or domination. While this makes space for a *social* critique of alienation at the level of its structuring causes,

³ Jaeggi (2013, 27 ff.) refers to Foucault in this context, but the idea is widespread across and beyond philosophical debates; see, for instance, Dennett and Kinsbourne (1992) and McConnel (2011).

no specific normative view about non-alienation needs to be endorsed in order to detect whether the relevant mechanisms trigger under those conditions.

2. Self-awareness and Self-understanding

The paradox of alienation is that alienated experiences are nonetheless our own: we are only alienated as *we* turn alien *to us*. Marcuse, however, famously argued that their contents can be experienced as true to the point that talks of alienation seem to become questionable “when individuals identify themselves with the existence which is imposed upon them and have in it their own development and satisfaction”, and yet this is nothing but “a more progressive stage of alienation”, one in which the subject is “swallowed up by its alienated existence” to the point that false consciousness seems to become true (Marcuse 1992, 13-14).

The problem Marcuse envisages is that false consciousness seems to be made true by agents that fully identify with their alienated existence. Marcuse understands alienation, along broadly Marxian lines, as an objective condition induced by exogenous social forces in which the attitudes, experiences, and actions a subject undergoes are not what they would have been, were those forces absent, and correlated with a subjective condition in which the subject’s awareness comes with a sense that they are not, or not fully, its own — a subjective experience of self-estrangement associated with a reduced or absent sense of agency, meaning, and affect (Cohen 1975, 120 ff.; Elster 1986, 77, 103-107; Wood 1981, 7-8). In this context, false consciousness is commonly designed to explain the fact that the self-understanding of agents typically mystifies alienation so as to reconcile them with their alienated existence and alleviate distress (Wood 1981, 13-14). According to Marcuse, however, in consumerist societies alienation reaches a perfected stage in which false consciousness seems to become true because subjects live through a way of life that embodies the ideology designed to mystify alienation

and thus identify with their alienated existence to the point of experiencing alienated attitudes, experiences, and actions as fully their own.⁴ In a sense, perfect alienation leaves no gap because agents are “swallowed up” by their alienated existence, thereby ceasing to perceive any sense of self-estrangement.

One does not need to endorse Marcuse’s claim entirely to appreciate the connection it captures between alienation and false consciousness. As a matter of fact, the distress commonly associated with alienation suggests that alienation is typically far from perfect and the related sense of powerlessness, meaninglessness, and indifference traces a distorted self-experience (Jaeggi 2014, 46-47). Yet Marcuse’s analysis casts light upon how alienation relates to false consciousness. Alienated attitudes, emotions, and actions are not attitudes, emotions, and actions we do not undergo; rather, they are attitudes, emotions, and actions we live through as alien. Thus, false consciousness cannot just amount to entertain a number of beliefs about our attitudes, emotions, and actions that fail to match the attitudes, emotions, and actions we undergo. Marcuse contends that false consciousness expresses the alienated self-awareness of people who fully identify with an alienated existence, which suggests that false consciousness must generally track an alienated self-awareness in order for the false beliefs it involves to yield a self-understanding that, at least to some extent, succeeds at hampering the capacity to perceive

⁴ Marcuse’s reading relies on the view, largely borrowed from Adorno, that in advanced industrial societies, ideology “is in the process of production itself” and therefore performs its function through products and services rather than by intentionally designed propaganda: mass transportation and communication; the industrial production of lodging, food, and clothing commodities; and the entertainment and information industry “indoctrinate and manipulate” simply because the things they deliver “carry with them prescribed attitudes and habits, certain intellectual and emotional reactions” that turn ideology into “a way of life”, thus shaping existence beyond belief. False consciousness becomes “immune against its falsehood” because ideology is “absorbed” into reality by being embodied in everyday artifacts, practices, and built environments (Marcuse 1992, 13-14).

one's own alienation. How can this connection between alienation and false consciousness be spelled out?

A self-understanding can be seen, roughly, as a set of *de se* ascriptions of attitudes, experiences, and actions, often arranged in a narrative form.⁵ Thus, *prima facie*, it might seem that self-understandings can be false like any predication, as they can turn out to attribute to their subject some property it fails to instantiate. As said, however, we cannot take false consciousness to be a set of false *de se* beliefs designed to cover alienation, because alienated attitudes, emotions, and actions are not attitudes, experiences, and actions we fail to go through. Rather, they are attitudes, emotions, and actions we live through as alien to us. Alienation looks paradoxical precisely because alienated subjects pervasively go through alienated attitudes, emotions, and actions. Therefore, in order to grasp how false consciousness relates to alienation, we need to focus on how self-understandings relate to our being aware of ourselves as subjects of the attitudes, experiences, and actions we undergo.

The point is that self-understandings are “self-referential stories” that are cast “in the mode of appropriation”: when we self-ascribe life episodes, experiences, and actions, we “identify with them, i. e. make them our own” (Crone 2020, 67). Telling a story as *my* story, as a narrative I identify with, however, is no part of the narrative itself. In fact, both the content and the narrative properties of a self-understanding — consistency, intelligibility, integration, and the like — would be preserved by a third personal report. Thus, a more basic sense of self must underlie appropriation, which tracks our first personal, pre-reflective self-awareness to be the subject of the relevant episodes, experiences, and actions (Zahavi 2007, 188-189, 191). In order to see how false consciousness relates to alienation, we need to focus on how it relates to disturbances affecting this minimal yet basic sense of self. The phenomenology reported in the

⁵ See Taylor (1995, 27 ff., 33-34, 47 ff.), Ferrara (1998, 29 ff., 162 ff.), McIntyre (1981, 116 ff.), Sandel (1988, 158 ff., 172), Korsgaard (1996, 83 ff.)

literature suggests indeed that in alienated experience the attitudes, experiences, and actions one goes through look like they are not, or not fully, one's own. It does not look like we self-ascribe, for instance, beliefs and desires we do not entertain. Rather, beliefs and desires we do entertain are experienced – at least to some extent – as if they do not belong to us. To resume Jaeggi's phrasing, they look like both belonging and not belonging to us. In order to track alienation, false consciousness must relate to a defective sense of being the subject of one's own attitudes, experiences, and actions.

In this connection, Marcuse depicts a situation in which, by participating in the "way of life" promoted by consumerist societies, agents appropriate a self-understanding that allows them to identify entirely with alienated attitudes, experiences, and actions so as to obliterate any sense of self-estrangement. We do not need to share this view of perfected alienation to realize that, in order to result in false consciousness, false *de se* ascriptions must track the alienated self-awareness of subjects who go through mental states and actions that look as if they are not their own. Appropriating a false self-understanding weakens the perception of alienation because it directs subjects towards perceiving alienated attitudes, experiences, and actions as genuine. The mistake false consciousness entails thus seems to involve a confusion concerning mental ownership. It does not consist in self-attributing attitudes, experiences, and actions that fail to match the attitudes, experiences, and actions one actually goes through, but in being wrong, or confused, about whether they are one's own.

This suggests that, insofar as they track alienated self-awareness, false *de se* ascriptions rest on an error through misidentification about the subject of self-attributed mental states and actions. The trouble with false consciousness draws on the feeling that this is an impossible mistake. Errors through misidentification diverge from ordinary predicative mistakes because they do not consist in predicating *F* of *a* when *a* is not *F*, but rather occur when one *knows* that *a* is *F* and judges that *b* is *F* on the ground of mistakenly believing that *a* is identical to *b*. For instance, I might know that Paul is tall and mistakenly believe that Mary is tall on the grounds

that I believe that Paul is identical to Mary. Now, it is commonly taken that *de se* ascriptions of mental predicates are immune to such errors, as it seems impossible for one to misidentify the subject of one's own attitudes, experiences, and actions. It seems to make no sense, for instance, to wonder whether the pain I feel is my pain (Evans 1982, 179–191, 215–225; Pryor 1999; Shumaker 1968; Wittgenstein 1968, 68; on actions, see O'Brien 2012, 141 ff.). A natural conjecture is that the use of "I" in this context is underpinned by our being pre-reflectively aware to be the subject of our own experiences, thoughts, and actions, which makes *de se* ascriptions of mental predicates immune to error through misidentification because of involving an indexical first personal acquaintance with their referent that is prior to reflective self-identification (Colombetti 2011; Morgan 2012; Perry 2002, 192 ff.; Zahavi 2014, 13 ff.). The appearance of a paradox concerning alienation arises indeed because in alienated experience we do not look wrong, or confused, about *what* we believe, desire, feel, or do, but about whether *we* are who believes, desires, feels, or acts. Relatedly, for a flawed self-understanding to track alienation, a disturbance affecting pre-reflective self-awareness must disrupt our sense of who is the subject of the attitudes, feelings, and actions we go through. False consciousness can only track alienated self-awareness insofar as a) some mechanism induces us to be wrong, or confused, about its subject and b) appropriating a false self-understanding prevents perceiving its alienated character. Jaeggi's cases are designed to convey vivid variations of this theme in everyday life: the pleased compliance with social roles of the oversocialized professional, the adventitious habits of suburban life, and the dissonant desires of the giggling feminist are "false" in that they are alien to the subject to which they are (self-)attributed (Jaeggi 2014, 52, 69-70, 101). The very possibility of false consciousness relies on the disruption of first personal, pre-reflective self-awareness that underlies the dynamics of alienation and appropriation.

The upshot is that, insofar as false consciousness involves identifying to some extent with a flawed self-understanding that tracks alienated self-awareness, it is not fully explained by well-known distorting mechanisms that affect attitude formation like confirmation bias, wishful

thinking, preference adaptation, evidence neglect, and so forth (see Elster 1983, 141 ff.; Elster 1986, 476 ff.; Mayerson 1991, Bianchin 2020, Bianchin 2021). These mechanisms might explain the content of false consciousness but do not bear on the distortion of pre-reflective self-awareness that is crucial to its subjective structure.

3. *Alienation and Mental Ownership*

A growing body of literature in philosophy and cognitive science contends that mental states and actions come with a pre-reflective sense of “mine-ness” that conveys mental ownership, which is a sense that we are who is living through the attitudes, feelings, and actions we undergo and differs from both the intentional and the attitudinal features of the latter. Although some non-trivial distinctions can be made among “mine-ness”, “for-me-ness”, and “me-ness”, it is common ground that mental ownership is an element of phenomenal consciousness that conveys pre-reflective self-awareness and allows *de se* ascriptions of mental predicates to express their subject (Guillot 2017; Lane 2012; Zahavi 2014, 18 ff.; Kriegel, Zahavi 2016; see also Damasio 1999, 168 ff.). My seeing a tree and remembering Paris differ in both attitude and content. Still, there is something common to them, namely their being *mine*, which differs phenomenally across persons. Your remembering Paris is not *mine*, even if our memories share the same content. In this context, we need not adjudicate how these different features of mentality connect. What is relevant is the claim that a sense of mental ownership must go along with experience, thought, and action. This typically involves three claims:

- a) Mental ownership accounts for *de se* ascriptions, as mental states and actions come with a phenomenal sense of mine-ness that makes them pre-reflectively accessible to their subjects (Zahavi 2005, 125-126; Kriegel, Zahavi 2016).

- b) Mental ownership displays a basic sense of self that anchors the extended sense of self subjects develop across time as an autobiographical self-understanding involving their social and cultural environment (Damasio 1999, 195 ff.).
- c) Mental ownership is phenomenally given and does not require the self to be free-standing with respect to conscious experience (Zahavi 2014, 20 ff.).

Thus, it seems that, insofar as false consciousness tracks alienated self-awareness, the only way to make sense of how consciousness can be false is thinking that mine-ness can fail (i.e., that our phenomenal pre-reflective sense of mental ownership can be led astray with the result that the mental states and actions we go through look like they are not ours). As we have seen, this looks like an impossible state of affairs if *de se* ascriptions are taken to be immune to error through misidentification because of resting on a first personal knowledge that is prior to reflective self-identification. In order for us to be possibly wrong, or confused, about their subject, there must be a way for pre-reflective mental ownership to be perturbed.

It might be tempting, at this point, to resort to affective experience to connect mental ownership and self-understandings so as to allow false consciousness to track alienation. We have seen that self-understandings need be appropriated in order for the values, beliefs, and roles they embody to play a role in our cognitive economy. Emotions, in this context, might be thought to allow detecting alienation because they carry information about ourselves as well as about the world by marking expected experiences and action outputs in evaluative terms.⁶ Thus, it might be tempting to think that self-understandings can be checked for whether they track

⁶ More specifically, the function of affective experiences can be seen as that of carrying information about the emotional significance of a bodily state produced by a context-sensitive appraisal of a situation, such as feeling sad differs from simply believing that a family member is dead because it carries information about the nature of our emotional response to the event (Gerrans 2019, 402, 407-409; De Sousa 1987, 182 ff.).

alienated self-awareness by considering whether our affective responses to actual or counterfactual situations are consistent with the way we understand ourselves (De Sousa 1987, 319 ff.; Salmela 2003, 310).

Taking affective experience as evidence, however, trivializes the point, as it only makes sense if we take for granted that emotions are necessarily non-alienated. This might seem right at first. While affective experiences can be wrong both with respect to what they are about and with respect to the kind of emotion they convey (De Sousa 2007, 330-331), it might seem that we cannot be wrong about who is being affected when they are self-ascribed. For instance, it might seem that we cannot be mistaken about who is afraid when we fall prey to fear. Yet in this respect affective experiences are on par with any other *de se* ascriptions. Thus, they provide no special evidence as to whether a self-understanding tracks an alienated self-awareness. Were *de se* ascriptions generally immune to error through misidentification, no *de se* ascription could possibly fail to pick up the right subject. Conversely, were *de se* ascriptions *not* immune to error through misidentification, there would no reason to think that affective experiences are safe from such error. Emotional alienation would indeed be a condition in which one feels alienated from one's own emotions, or, more precisely, a condition in which emotions appear to be both owned and disowned by the same agent (Szanto 2017, 266).

In fact, clinical evidence suggests that mental ownership can be disrupted, and therefore *de se* ascriptions are vulnerable to error through misidentification (Klein 2015). In particular, the literature about depersonalization suggests that we can be aware of our mental states and actions without it seeming to us that they are our own. Depersonalization designates a condition often associated with severe depression, schizophrenia, and anxiety disorders, as well as with trauma-related dissociation, in which patients report feeling detached from their bodies, mental states, and actions, suggesting a disruption in their phenomenal sense of mine-ness (Klein 2015, 366). Similarly, patients suffering from Cotard's syndrome are under the delusion of not having bodily organs or mental states, and in severe cases they deny being alive, which has been read

as resulting from sensing that their bodily organs do not feel as their own, that thoughts and emotions feel alien, and that one “is not an ‘I’ anymore” (Billon 2016, 357, 375-376). These cases convey evidence that mental ownership can be disrupted and that a variety of mental disorders underpins the disruption. Milder and transient forms of depersonalization, however, can be elicited by sleep deprivation and recreational drugs, which suggests that the disruption varies in intensity and extension, ranging from mild, transient experiences in which mental states do not clearly feel as one’s own to acute cases in which patients feel like they do not exist (Billon 2014, 745 ff.).

Given these differences, a sensible hypothesis is that mental ownership is underpinned by a specific mechanism whose functioning can be affected by incumbent psychopathologies or traumatic events that alter the normal conditions under which it is designed to work (Klein 2015, 369). For instance, Gerrans (2019) suggests that the absence or aberrance of sensory feedback about emotions, perceptions, and action outputs, which is typically related to mental disorders or traumatic events, might induce the organism, given the otherwise intact working of cognition and perception, to feel as if it is no longer present in experience. However, the fact that disturbances in the phenomenology of mental ownership are gradable and that milder cases can be elicited by causal factors that do not involve severe mental disorders suggests that the relevant mechanism is also vulnerable to malfunctions that are not underpinned by severe psychopathology. This is relevant to understanding alienation. Alienation apparently involves a disturbance in the phenomenology of mental ownership unrelated to severe mental disorders and underpinned by exogenous, social factors rather than by endogenous psychopathological conditions. On the one hand, alienation looks close to mild depersonalization, which makes it compatible with mental health. On the other hand, unlike transient mild cases underpinned by occasionally abnormal individual conditions like sleep deprivation or drug use, alienation is by hypothesis underpinned by non-occasional social arrangements like the commodification of labor or the structural injustice engendered by race- and gender-related social norms, which

makes it enduring and resilient to psychological intervention.⁷ A natural conjecture is that social facts can affect mental ownership by perturbing the normal conditions under which it is designed to work.

A well-known distinction between triggering and structuring causes can be used at this point to disentangle alienation from clinical depersonalization. Triggering causes explain what in a causal process caused *C* which caused *E*; structuring causes are designed to explain what “shaped or structured the process” — what made *C* to cause *E* rather than something else (Dretske 1988, 42). This distinction has been recently used to make sense of social structural explanations (Haslanger 2016, Soon 2021). In the present context, it allows to sort out alienation from clinical depersonalization by discerning the distinctively social dimension of the former. As seen, clinical depersonalization is associated with severe mental disorder, which means that incumbent psychopathologies and traumatic events work as structuring causes that perturb the normal conditions under which the mechanism for mental ownership works so as to generate abnormal sensory feedback that triggers it to malfunction. Thus, clinical depersonalization is typically severe, persistent, and resilient. Alienation differs in that social factors seem to work in this case as structuring causes of a milder, yet non-transient condition of depersonalization. For instance, the social roles one occupies, the socio-economic positions they are associated

⁷ Alienation is typically taken to be an enduring condition that causally depends on social structural features of the social environment like, eminently, production relations under capitalism (Evans 2022: 132 ff.). Depersonalization is therefore likely to be milder than in clinical cases, but nonetheless resilient to individual treatment. A parallel with socially induced cognitive distortions like implicit bias can be enlightening. Debiasing procedures in which individuals are trained to associate non-stereotypical traits to certain groups can succeed for the time being, but individuals tend to resume their biased habits as soon as they return to the original social environment (Liao, Huebner (2021: 105). Similarly, while it seems sensible to acknowledge that the experience of meaninglessness, agential deprivation, and emotional indifference associated with alienation can be temporarily alleviated, changes are likely to be volatile as long as social facts are unchanged.

with, and the power relations they entail are routinely taken to underpin evidence neglect and dispose people to assimilate biased information, which in turn affect the sensory feedback about one's own emotions, perceptions, and actions (e.g., modulating how an emotion feels, how we perceive interpersonal relations, or how our actions affect the social environment).

Focusing on structuring causes thus allows for a social structural explanation of alienation that steers clear from reducing alienation to a purely psychological condition by modeling how depersonalization can be socially induced. For instance, social pressure to comply with gender-, class-, or race-related norms can be taken to induce biased interpretation of actions and events that alter our sensitivity to what we perceive, feel, or do. Marx's case for the alienating effects of wage labor is perhaps less obvious, but even more revealing of how social structuring causes work. Marx takes alienation to be an objective relationship between agents and a social state of affairs that correlates with a distorted relationship between agents and their mental states and actions – a subjective condition of self-estrangement. That workers surrender their labor to capitalists is an objective social fact that does not depend on their attitudes towards labor but on the commodification of labor itself. That they fail to recognize themselves in their labor activities and products, feel reified into productive tools, and are detached from their social and natural environment, is a subjective condition that depends on the reduced sense of agency that goes along with not being in control of the production, being expropriated of its products, and being placed in a social bond “expressed in exchange value”, where “the social connection between persons is transformed into a social relation between things” (Marx 1993 [1857-8], 156-157; Forst 2017, 39 ff.; Lukes 1967; Tiffany 2003, 415-416). On the conjecture sketched above, the connection between the objective and subjective dimensions of alienation is explained by casting wage labor, by which workers are objectively alienated, as a structuring cause of their subjective alienation. Workers are objectively alienated because capitalism turns labor into a commodity, and they are subjectively alienated because they consequently fail to perceive their labor activities as their own actions and to recognize themselves in their products.

4. Simulations

In the previous section, I argued that the disturbances of mental ownership involved in alienation must be weaker than clinical depersonalization and prone to trigger under social rather than psychopathological structuring conditions. In this section, I suggest that some of the mechanisms underlying social cognition mediate the process by allowing subjects to go through alien mental states and actions *as if* they are their own. Take someone who appropriates a spurious self-understanding: a wannabe *maudit*, a *bourgeois bohemién*, or a revolutionary poser bringing themselves to think, feel, and act as if they were their imaginary counterparts. They go through beliefs, desires, emotions, and actions that, in a sense, are not their own. The latter are rather simulations of their counterpart's genuine beliefs, desires, feelings, and actions. Over time, however, simulated states and actions become entrenched habits, and the memory of simulation recede to the point that it is no longer clear whether they are genuine or spurious. Still, a few, perhaps barely visible traces betray their origins: as in Sartre's *mauvais foix*, actions look too precise, expressions slightly mechanical and stereotyped, behavior somehow rigid (Sartre 1978, 59-60). Subjects possibly feel like they are the persona they modeled themselves on, yet a disturbing feeling occasionally surfaces of being detached from what they think, feel, and do.

One would contend that they are alienated, were it not for the fact that they intentionally managed themselves into living through alien attitudes, experiences, and actions. It is perhaps fair to say that their alienation is harmless because the causal process is not structured by endogenous pathological conditions or exogenous social factors. How about unintentional cases, in which agents do not intend to alienate themselves and simulation routines are beyond their control, escape awareness, and are induced by exogenous social factors? Simulation

routines need not be intentional, or conscious, and can be induced by external as much as by internal forces, such as by the pressure to conform to widely accepted social norms or to acquiesce in socio-economic objective conditions that leave no better alternative than to adapt. Gender norms and the labor market are prime examples in the literature. This would fit Jaeggi's case for alienation as a condition in which subjects experience the beliefs, desires, feelings, or actions they go through as if they are not, or not fully, their own — as if they both belong and do not belong to themselves.

While the view that we can put ourselves in the mental shoes of others and feel *as if* we were there is far from new, the relevant mechanisms for social cognition have been recently outlined in some detail. Theories of social cognition notoriously come in a wide variety, and current research leans toward a pluralistic approach (Andrews et al. 2020). In what follows, I focus on a cluster of rather established ideas that marks off simulation theories and suggest that they contribute to account for alienation and its relation to false consciousness. It is a matter of further investigation whether and how other approaches can be recruited to this effect. Simulation theories are traditionally contrasted with theories of mind. While the latter take social cognition to rely on a native or early acquired theory of mind designed to explain behavior by positing mental states and processes as internal causes, the former take it to rely on simulating them by displacing in the mental shoes of others and running our own cognitive mechanism off-line. The simplest forms of simulation are often taken to consist in a neural process that mirrors the neural process underlying observed actions and to rely on a “mirror system” that matches the execution and the observation of actions (Gallese and Goldman 1998; see Sinigaglia and Rizzolati 2006). Mirror neurons, however, only operate on observed actions and only read off intentions-in-action, which leaves out prior intentions as well as any mental state or process that is not embodied in observed behavior (Jacob and Jannerod 2005; Goldman 2013, 100-101).

In order to reach beyond observed actions, simulation relies on imagination (Goldman 2006, 147 ff.). Gordon (1995a) argues in particular that simulative mindreading involves an “imaginary shifting in the reference of indexicals” that alters the reference of indexical expressions like “I”, “here”, and “now” so as to recenter the egocentric maps associated with first personal drives to action and emotion. The background idea is that information only affects our cognitive economy by being connected with irreducibly indexical first personal knowledge locating the agent: what triggers action and emotion is “[...] the lion coming after *me*, the meeting *I* am supposed to be *now*, the insult directed to *me*, the award given to *my* child” (Gordon 1995a, 773-774). Thus, shifting indexicals change not only egocentric spatial and temporal relations, but also how personal and social relations are mapped. On this reading, theoretical inference plays no role, as there is no need to “mentally transfer a mental state from one person to another”: all we need to do is let our own cognitive processes run as if we were there (Gordon 1995b, 56). Gordon’s prototypical example comes from acting practices that require actors to be motivated “from within” to perform what the enacted persona says and does, which involves “transforming themselves” to become the character they play (Gordon 1995a, 735).

Simulation is designed to underpin cultural learning and the creation of social bonds as well as small- to large-scale cooperation and socialization processes (Goldman 2006, 276 ff., Tomasello 2014, Bianchin 2015). A number of operations, however, is performed in the way. First, pretend mental states are created in imagination that are approximate equivalents of genuine first-order mental state and they are attributed to others as *input* to deliberation. Second, a deliberative process is run to generate a decision. Finally, an action is predicted. Simulation is thus driven by processes occurring in an interpreter who uses herself as a model rather than building a theoretical model of the target (although it can exploit tacit theories), which implies that occurrent states and processes in both systems are of the same kind and level (Goldman 2006, 34 ff.). In this context, imagination does not consist in *a sui generis* propositional attitude

whose contents represent a counterfactual state of affairs, but as a capacity to enact counterfactual mental states, processes, and actions. Understood as a propositional attitude, imagination is a metacognitive state that differs in kind and order from the mental states, processes, or actions it takes as objects. On this reading, imagining *that* Paolo desires that P does not involve entertaining the desire that P, but entertaining a higher-order attitude of a different kind — imagination — whose content represents the fact that Paolo desires that P. On the enactive reading, instead, imagination does not represent but enacts the target states, processes, or action: imagining Paolo's desire that P involves entertaining the desire that P. It does not represent *that* Paolo desires to swim, but feels like Paolo's desire to swim. Simulative mindreading involves the latter “enactment imagination” (Goldman 2006, 46-48, 195 ff., 281 ff.).

Simulation theory thus allows for a literal sense in which we can undergo mental states and processes that are, in a sense, not our own. Accounting for alienation, however, requires simulations to escape consciousness and intentional control to the point that simulated states, processes, and actions are conflated with our own and therefore experienced at least to some extent as alien states, processes, and actions inhabiting our minds. This means that our awareness and control of the related simulation must be inhibited either completely or partially.

We can conjecture which parts of simulation routines can be distorted to this effect. First, simulation requires first personal access to one's own mental states and the capacity to classify them, but there is no reason to think that mechanisms underlying first personal access and classification are less fallible than others (Goldman 2006, 148). Thus, mental states can be misclassified and pretended states be confused with genuine states. Second, simulative mindreading is reliable only if projecting egocentric biases is prevented, which requires a quarantine mechanism to seclude biased attitudes from those we are allowed to attribute to others (Goldman 2006, 40-41). This is meant to rule out projection rather than alienation, yet again suggests that the input undergoes selection operations that escape awareness and is prone

to manipulation. Third, as said, the initial pretend states are approximate equivalents of genuine first-order states and differ from them only with respect to their origin, which could be forgotten as cognition proceeds. Thus, simulation requires that self-attributed mental states are kept apart from the mental states attributed to others. Hence, simulating systems must be endowed with the capacity to “tag” simulated mental states as being owned by somebody else in order to retain the distinction across time (Goldman 2006, 28, 30, 186-187, 211-212). Again, there is no reason to think that this capacity is infallible, and there is some evidence that disturbances in the relevant brain regions result in its disruption (Goldman 2006, 212-213, 153 ff., 161 ff.). As a consequence, attributed states and processes can get confused with one’s own genuine states and processes over time.

To put it differently, the mechanisms that govern attribution can fail, so that the outputs of simulation get self-attributed as one’s own mental states as well as vice-versa. Schizophrenic patients, for instance, are reported to attribute to others comments and orders they address to themselves in the third person, and generally to experience a weakened sense of ownership and a reduced self–other demarcation (Froese and Krueger 2020; Jannerod and Pacherie 2004). Alienation seems to entail a germane malfunction inducing subjects to confound attitudes, emotions, and actions that are implicitly attributed to others with one’s own, which results in feeling like entertaining alien thoughts, intentions, and emotions — in feeling them as both belonging and not belonging to oneself. This suggests that mental ownership can be affected when simulation routines are disguised, neglected, or forgotten. As Goldman suggests, genuine and spurious states can get confused so that agents go through simulated attitudes, emotions, and actions and receive correspondingly reduced sensory feedback about them, yet fail to tag them as attributed to others. In such cases, one can expect agents to feel like they are going through actions over which they have no full control, emotions that feel alien, and attitudes that do not look clearly their own.

This fits some major characterizations of alienation. Karen Horney (1945), for instance, takes alienation as the condition of someone who develops an “idealized image” of himself and “cling[s] to the belief” that he is his idealized image to the point of becoming “oblivious to what he really feels, likes, rejects, believes – in short to what he really is”, which results in a “paralysis of initiative and action” and a “permanent condition of being unreal to oneself” (110-111). The idea that identifying with the idealized image is a matter of belief needs to be revised, yet once this is replaced by a shift in one’s own egocentric maps along the lines of simulation theory, the resulting self-obliviousness, practical inertia, and emotional detachment can be explained by taking the identification to result from obliterating the simulation process. A virtue of this reading is that it does not require committing to any substantial conception of the “real” self and that it leaves space for the identification process to be possibly less than perfect. Still, Marcuse’s view of perfected alienation could be perhaps vindicated if it turned out that the simulation process can be *completely* obliterated under the favorable conditions of affluent consumerist societies in which subjects are “swallowed up” by the one-dimensional way of life which is imposed upon them and therefore fully identify with their alienated existence. False consciousness looks true because alienation leaves no gap. Conversely, Jaeggi’s suggestion to the contrary would be read as intimating that some traces of simulation, however blurry or ambivalent, are likely to persist and interfere with full identification. The suburban academic, the over-socialized professional, the giggling feminist, as well as the disillusioned professor are not fully unaware that something resists identification, which explains their persistent experience of self-estrangement. On this reading, alienation can come in degrees and typically results in the constellation of muteness, lack of control, and indifference traditionally associated with alienated experience. Accordingly, false consciousness is far from being as “immune to its falsehood” as Marcuse thought (Marcuse 1992, 14).

5. Conclusion: Social Structural Explanations and Alienation Critique

In this paper, I have argued that talks of alienation make sense because social factors can unsettle the conditions under which our sense of mental ownership normally works, have shown how alienation relates to false consciousness, and have outlined a social structural explanation of how social factors can affect the mechanism underlying mental ownership by affecting the simulation routines through which we are socialized into groups and larger social arrangements. According to this view, learning social roles, internalizing the related norms and values, and adjusting to the social position we occupy can be conducive to alienation and yield false consciousness. Admittedly, simulation can play a harmless or a mixed role. Religious conversion, for instance, is likely to involve a desire for transformation and a process of self-construction through which agents model themselves on exemplary characters and use simulation to explore and guide future conduct. Eventually, the converted are likely to feel transformed, and it is hard to take the process to be unquestionably one of self-alienation, although borderline cases where genuine motives intermingle with more ambivalent drives might be common. Things look differently, however, where indoctrination, ideology, or forced adaptation to social roles and position is involved. Consider someone who finally adjusts to social conditions she cannot escape and identifies with social roles associated to her socio-economic position, be it only to reduce cognitive dissonance, or consider the mimetic behavior of a professional motivated by peer pressure to comply with social norms (Mallon 2013, 162 ff.). Over time, the enacted roles may slowly worm into the agent and take control as the memory of simulation recedes and compliance becomes habitual. These cases seem likely to yield the sort of mild depersonalization that marks off alienation. Also, consider education. It seems wrong to conclude that education alienates just because training in a practice largely depends on imitation. Yet things look different if we focus on how class, race, and gender roles are learned by modeling oneself on exemplary figures who routinely embody biased stereotypes

and ideological representations of class, race, and gender. It seems safe to say that talks of alienation are legitimate when social factors induce simulation routines that are forced and opaque to their subject because the disposition to identify with social positions or roles flows from structural forms of domination or injustice associated, for instance, with capitalist market structures (Vrousalis 2021), gender (Haslanger 2012, 8, 314 ff.), or structural racism (Shelby 2016, 28 ff.).

Whatever the normative conception of non-alienation, disturbances affecting mental ownership under such conditions can be regarded as paradigmatic cases of alienation. On the present reading, they are due to the fact that the mechanisms underlying mental ownership are brought to trigger under socially induced aberrant conditions that alter our sensitivity to what we think, feel, and do. Locating the mechanisms at work allows talks of alienation to be more than metaphoric, while connecting their malfunction with exogenous social factors suggests that focusing on structural features of social arrangements is crucial to single out alienation as a socially induced condition.

Two implications are important. First, alienation here is subject to a social structural explanation rather than to a purely psychological account. Second, overcoming alienation consequently involves social critique and transformation rather than psychological diagnosis and therapy. Incidentally, this allows to steer clear from an objection that has been raised against Jaeggi's reading, namely that it cannot locate specific social causes for the agents' failure to appropriate the form of life they inhabit, so that the overcoming of alienation turns out to look as a mere psychological change in their attitudes and dispositions (Evans 2022, 132-133). Whatever the merit of this objection, the demand for social causes Jaeggi's reading is allegedly found missing is explicitly met here by explaining the rather elusive sort of self-estrangement alienation entails as a disruption in the sense of self whose structuring causes are located at the social rather than at the individual level. On this reading, alienation can be detected wherever self-estrangement can be causally traced back to conducive social conditions rather than to

underlying psychopathological conditions. Demarcating alienation from clinical depersonalization by its etiological profile accordingly locates the space for alienation critique at the social level of its structuring causes.

While not itself a normative consideration, this places a normative demand on the theory of alienation that is different from those commonly involved in conceptualizing non-alienation as the standard of alienation critique, because it rather concerns the social arrangements under which alienation materializes. Detecting a malfunction in the mechanisms that underlay mental ownership where no mental disorder or other psychologically abnormal conditions are found working as structuring causes invites the conjecture that there must be something wrong with the social conditions under which those mechanisms are brought to malfunction. To the extent that the flaws from which alienation originates are located at the social level of its structuring causes, alienation critique should hence be expected to emerge from explaining how alienation comes about. In this sense, the normative task of a theory of alienation cannot be divorced from its explanatory task.

References

- Andrews, K., Spaulding, S., Westra, E. (2020). Introduction to Folk Psychology: Pluralistic Approaches. *Synthese* (forthcoming).
- Bianchin, M. (2021). Ideology, Critique, and Social Structures. *Critical Horizons* 22(2): 184-196.
- Bianchin, M. (2020). Explaining Ideology: Mechanisms and Metaphysics. *Philosophy of the Social Sciences* 50(4): 313-337.

- Bianchin M. (2015). From Joint Attention to Communicative Action: Some Remarks on Social Ontology, Critical Theory, and Cognitive Science. *Philosophy & Social Criticism* 41(6): 593-608.
- Billon, A. (2014). Why Are We Certain that We Exist?. *Philosophy and Phenomenological Research* 91(3): 723-759.
- Billon, A. (2016). Making Sense of the Cotard Syndrome: Insights from the Study of Depersonalisation. In *Mind & Language* 31(3): 356-391.
- Ciaunica, A., Charlton, J., Farmer, H. (2020), When the Window Cracks: Transparency and the Fractured Self in Depersonalisation. *Phenomenology and the Cognitive Sciences* 20(1): 1-19.
- Ciaunica, A. Pienkos, E. Nakul, E. Madeira, L., Farmer, H. (2022). Exploration of self- and world experiences in depersonalization traits. *Philosophical Psychology* (forthcoming). DOI: 10.1080/09515089.2022.2056009
- Cohen, J. (1978). *Marx's Theory of History*. Oxford: Oxford University Press.
- Colombetti. G. (2011). Varieties of pre-reflective self-awareness: foreground and Background bodily feelings in emotion experience. *Inquiry* 54(3), pp. 293-131.
- Crone, K. (2020). The self-understanding of persons beyond narrativity. *Philosophical Explorations. An International Journal for the Philosophy of Mind and Action* 23(1): 65-77.
- Damasio, A (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.
- Davidson, D. (1985). *Deception and Division*. In *The multiple self*, edited by Jon Elster, 79-92. Cambridge: Cambridge University Press.
- De Sousa, R. (1987). *The Rationality of Emotions*. Cambridge (MA): MIT Press.
- De Sousa, R. (2007). Truth, authenticity, and rationality. *Dialectica* 56(2): 323-345.

- Dennett, D., Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioral and the Brain Sciences* 1(2): 183-201.
- Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. Cambridge (MA): MIT Press.
- Elster, J. (1983). *Sour Grapes; Studies in the Subversion of Rationality*, Cambridge: Cambridge University Press.
- Elster, J. (1986a). *Making Sense of Marx*. Cambridge: Cambridge University Press.
- Elster, J. (1986b). *An Introduction to Karl Marx*. Cambridge: Cambridge University Press.
- Evans, G. (1982). *Varieties of Reference*, Oxford: Oxford University Press.
- Evans, J. (2022). Rahel Jaeggi's Theory of Alienation. *History of the Human Sciences* 35(2): 126-143.
- Ferrara, A. (1998). *Reflective Authenticity: Rethinking the Project of Modernity*. London: Routledge 1998.
- Forst, R. (2017). Noumenal Alienation: Rousseau, Kant and Marx on the Dialectics of Self-Determination. *Kantian Review* 22(4): 523-551.
- Froese, T., Krueger, J. (2020). Lost in the Socially Extended Mind: Genuine Intersubjectivity and Disturbed Self-Other Demarcation in Schizophrenia. In *Time and Body: Phenomenological and Psychopathological Approaches*, edited by C. Christian Tewes and Giovanni Stanghellini, 318-340. Cambridge: Cambridge University Press.
- Galeotti, E. (2018). *Political Self-Deception*. Cambridge: Cambridge University Press.
- Gallese, V., Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2(12): 493-501.
- Gerrans, P. (2019). Depersonalization Disorder, Affective Processing and Predictive Coding. *Review of Philosophy and Psychology* 10(2): 401-418 407-409.
- Goldman, A. (2013). *Joint Ventures: Mindreading, Mirroring, and Embodied Cognition*. Oxford: Oxford University Press.

- Goldman, A. (2006). *Simulating Minds. The Philosophy, Psychology and Neuroscience of Mindreading*. Oxford University Press.
- Gordon, R. (1995a). Sympathy, Simulation and the Impartial Spectator. *Ethics* (105) 4: 727-742.
- Gordon, R. (1995b). *Simulation without introspection or inference from me to you*, in M. Davies, T. Stone (a cura di), *Mental simulation* (pp. 53-67). Oxford: Blackwell, 1955, p. 56.
- Guillot, M. (2017). I Me Mine: on a Confusion Concerning the Subjective Character of Experience. *Review of Philosophy and Psychology* 8(1): 23-53,
- Haslanger, S. (2012). *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press.
- Haslanger, S. (2016). What is a (Social) Structural Explanation? *Philosophical Studies* 173(1): 113-130.
- Honneth, A. (2012). *Reification*, Oxford: Oxford University Press.
- Horney, K. (1945). *Our inner conflicts*. New York: Norton.
- Jacob, P, Jannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in cognitive science* 9(1), 21-25.
- Jaeggi, R. (2014). *Alienation*. New York: Columbia University Press.
- Jaeggi, R. (2017I. Pathologies of Work. *Woman's Studies Quarterly* 45(3-4): 59-76.
- Jannerod, M., Pacherie, E. (2004). Agency, Simulation, and Self-Identification. *Mind and Language* 19(2): 113-146.
- Klein, S. (2015), The Feeling of Personal Ownership of One's Mental States: A Conceptual Argument and Empirical Evidence for an Essential, but Underappreciated, Mechanism of Mind. *Psychology of Consciousness: Theory, Research, and Practice* 2(4): 355-376.
- Korsgaard, K. (1996). *Sources of Normativity*. Cambridge: Cambridge University Press.
- Lane, T. (2012). Toward an explanatory framework for mental ownership. *Phenomenology and the Cognitive Science* 11(2): 251-286.

- Liao, S., Huebner, B. (2021). Oppressive Things. *Philosophy and Phenomenological Research* 103(1): 92-113.
- Lukes, S. (1967). Alienation and Anomie. In S. Lukes (1977). *Essays in social theory*, 74-95. New York: Columbia University Press.
- Mallon, R. (2013). *The Construction of Human Kinds*. Oxford: Oxford University Press.
- Marcuse, H. (1992). *One-Dimensional Man*. London: Routledge.
- Marx, K. (1993[1857-8]) *Grundrisse*. London: Penguin
- Mayerson, D. (1991). *False Consciousness*, Oxford: Oxford University Press.
- McConnell, R. A. (2011). The Multiple Self-Aspects Framework: Self-Concept Representation and Its Implications. in *Personality and Social Psychology Review* 15(1) 3–27.
- McIntyre, A. (1981). *After Virtue: A Study in Moral Theory*. Notre Dame: Notre Dame University Press.
- Morgan, D. (2012). Immunity to error through misidentification: what does it tell us about the *de se* ? in *Immunity to Error Through Misidentification: New Essays*, edited by Simon Prossner and François Recanati, 104-123. Cambridge: Cambridge University Press.
- O'Brien, L. (2012). Action and Immunity to Error Through Misidentification. In *Immunity To Error Through Misidentification: New Essays*, 124-143. Cambridge: Cambridge University Press.
- Perry, J. (2002). *Identity, Personal Identity, and the Self*, Hackett, Indianapolis 2002.
- Pryor, J. (1999). Immunity to Error from Misidentification. *Philosophical Topics* 26(1/2): 271-304.
- Rizzolati, G., Sinigaglia, C. (2006). *Mirrors in the Brain: How Our Minds Share Actions and Emotions*. Oxford: Oxford University Press.
- Rosa, H. (2013). *Social Acceleration: A New Theory of Modernity*. New York, Columbia University Press.

- Rosa, H. (2019). *Resonance: A Sociology of Our Relations with the World*. Oxford: Polity Press.
- Salmela, M. (2003). What is Emotional Authenticity? *Journal for the Theory of Social Behavior* 35(3): 309-330.
- Sandel, M. (1998). *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press.
- Sartre, J.P. (1978). *Being and Nothingness*. New York: Pocket Books.
- Shelby, T. (2016). *Dark Ghettos: Injustice, Dissent, and Reform*. Cambridge (MA): Harvard University Press.
- Shumaker, S. (1968). Self-Reference and Self-Awareness. *The Journal of Philosophy* 65(19): 555-567.
- Sierra, M. (2009). *Depersonalization: A New Look at a Neglected Syndrome*. Cambridge: Cambridge University Press.
- Sierra, M., & David, A. S. (2011). Depersonalization: A selective impairment of self-awareness. *Consciousness and Cognition*, 20(1), 99–108.
- Soon, V. (2021). Social Structural Explanation. *Philosophy Compass* 16(10): e12783-e12799.
- Szanto, T. (2017). Emotional Self-Alienation. *Midwest Studies in Philosophy* 41(1): 260-286.
- Taylor, C. (1995). *Sources of the Self: The Making of Modern Identity*. Cambridge: Cambridge University Press.
- Tiffany, E. (2003). Alienation and Internal Reasons for Action. *Social Theory and Practice* 29(3): 387-418.
- Tomasello, M. (2014). *A Natural History of Human Thinking*. Cambridge (MA): Harvard University Press.
- Vrousalis, N. (2021). The Capitalist Cage: Structural Domination and Collective Agency in the Market. *Journal of Applied Philosophy* 38 (1):40-54.
- Wittgenstein, L. (1968). *The blue and brown books*. New York, NY: Harper and Row 1968.

Wood, A. (1981). *Marx*. London: Routledge.

Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First-Person Perspective*.

Cambridge (MA): MIT Press.

Zahavi, D. (2007). Self and Other: The Limits of Narrative Understanding. *Royal Institute of Philosophy Supplement* 60: 179-201.

Zahavi, D. (2014). *Self and Other: Exploring Subjectivity, Empathy, and Shame*. Oxford: Oxford University Press 2005.

Kriegel, U., Zahavi, D. (2016). For-Me-Ness: What It Is and What Is Not. In *Philosophy of Mind and Phenomenology Conceptual and Empirical Approaches*, edited by Daniel O. Dahlstrom, Andreas Elpidorou, and Walter Hopp, 36-53. London: Routledge.