

This is a preprint version of the paper that appeared in the *Journal of Philosophy*. Please cite the printed version.

“Truthfulness and sense-making: two modes of respect for agency” *The Journal of Philosophy*, 121 (2), 61-88, 2024.

TRUTHFULNESS AND SENSE-MAKING: TWO MODES OF RESPECT FOR AGENCY*

What are the requirements of respect for others *as agents*?¹

Chris regularly travels to another town for work and sometimes attends conferences further afield. When at conferences Chris often engages in casual sexual liaisons and is also conducting an on-again, off-again affair in the town he travels to for work. Chris has a spouse, Jo, and two teenage children. Chris and Jo get on well; their children are navigating adolescence with mixed success. Chris is comfortable with his extra-curricular activities and judges it best not to tell Jo about them. Jo would feel betrayed, and might even end the marriage which would likely be bad for their children and disruptive all round. So Chris lies as necessary, and endorses Jo’s ideal conception of their marriage in their friendship group.

Chris may be right that, *all things considered*, it would be better to keep Jo in the dark. It is possible that everyone will be happier if the status quo is maintained. We do not wish to adjudicate that point here. Even so, it appears that Chris wrongs Jo. Jo is being deliberately deprived of information that is clearly material to her values, her interests, and to decisions that she might make regarding her marriage. Chris thus substitutes his own judgment for Jo’s. She is not given the opportunity to reflect on the true state of their marriage or a voice in the decisions Chris makes about it.

The wrong of deception, has, since Kant, been characterised as a failure of respect for the agency of the other. According to Kant, as rational beings capable of evaluating and setting their own ends, persons are not to be treated as mere means to another’s end. In lying we manipulate the other’s rational capacities in order to achieve ends we know, or fear, they would not share. This is paradigmatically a failure of respect. Truthfulness then, is seen as a central mode of respect for each other’s agency.

We agree. But we claim that a closer examination of our goals *qua agents* and of the ways in which agency can be supported or undermined in our interactions with each other reveals a further and distinct mode of respect for agency. The importance of truthfulness lies in significant part in the ways in which it answers to and supports our agential need to

* The authors wish to thank two anonymous referees for this JOURNAL for their extensive remarks responding to earlier versions of the manuscript. Both reviewers’ insightful critical engagement greatly improved the article. We also especially thank Doug McConnell for generously reading an early draft and for written comments. For helpful discussions we thank audiences at the University of Melbourne Philosophy Department, and the Philosophy department at Deakin University.

¹ We are primarily concerned in this paper with what Stephen Darwall has termed *recognition respect*. See Stephen Darwall, “Two Kinds of Respect,” *Ethics*, LXXXVIII, 1 (1977): 36–49. For clarity, we deliberately focus on agency rather than autonomy since we are concerned broadly with duties of respect towards persons who are undoubtedly agents but who may or may not enjoy autonomy or who experience (as we all do) varying levels of autonomy throughout their lives. While much of what we say may connect to discussions of autonomy, agency is the more fundamental, and, we think, the more important notion.

make intelligible, to *make sense of* our world, other people, and ourselves. Thus, we take as the main burden of this paper to draw out the notion of sense-making and highlight its importance for human agency. Since sense-making is something we often do together, and that we can support or undermine, it generates norms of interaction that we claim constitute a distinctive mode of recognition and respect for another's agency.

Truthfulness and support for sense-making are both modes of respect for agency, however our analysis exposes a faultline between them. We describe a variety of situations where a rigid insistence on the truth disrupts sense-making and risks disrespect for the agency of the other. What should we do when the requirements of truthfulness conflict with support for sense-making?²

The paper will proceed as follows. We first utilize J. David Velleman's theory of agency to describe the concept of sense-making, its link to self-image, and its centrality in diachronic agency. We then outline the ways in which we facilitate each other's sense-making in interpersonal relationships and show how undermining sense-making is a serious failure of the respect we owe to each other. We trace the importance of truthfulness for the sense-making project, and, through a series of cases, provide an analysis of why and when divergences from being wholly truthful might be morally justified on grounds of respect for agency.

We close with a discussion of the limit case of dementia to develop the claim that in some cases of marginal agency the importance of sense-making wholly overrides the norms of truthfulness, and to defend the notion that our sense-making needs ground a requirement of respect even in such cases of diminished agency. We then defend our claim against two objections. First, that the self-understanding of those in the marginal cases is insufficiently robust for sense-making to occur, and second, that it is beneficence, not respect, that justifies departures from truthfulness in such difficult cases.

I. SENSE-MAKING, SELF-IMAGE AND DIACHRONIC AGENCY

According to Velleman,

To act for a reason is to do what would make sense, where the consideration in light of which it would make sense is the reason for acting...When one's behaviour is guided by such considerations, it is guided by one's capacity for making sense of behaviour...³

At the most basic level the role of sense-making in agency is revealed in our need to comprehend what we are doing at any given moment. This understanding can be implicit, so it is consistent with our doing something intentional without thinking about it, such as when we are driving a car while thinking about something else. But if we literally lose our grip on

² Following Sissela Bok we distinguish *truthfulness* from *truth-telling*, in favour of the former. See Sissela Bok, *Lying: Moral Choice in Public and Private Life* (New York: Random House, 1978). A person can select certain truths they intend to convey while thereby effectively deceiving their interlocutor. By 'truthfulness' we will mean intended truth-telling that is sufficiently complete and relevant, as to avoid deceitfulness. This wide definition incorporates honesty as its central norm, though a cluster of related communicative intentions are involved, such as sincerity, candidness, being forthright, frank, and so on. Our definition echoes that of Bernard Williams who emphasised two truthfulness virtues, that of accuracy, and that of sincerity; see Bernard Williams, *Truth and Truthfulness: An Essay in Genealogy*, (Princeton: Princeton University Press, 2002).

³ David Velleman, *Self to Self* (Cambridge: Cambridge University Press, 2006), p.6.

what we are doing our agency is interrupted and we may cease, temporarily, to act.⁴ Velleman invites us to reflect on the familiar experience of forgetting what we are doing.⁵ We walk into a room purposefully but then forget what we came in for. What do we do? As Velleman says, we stop. “What am I doing in here?” we think. We cannot do anything until we figure it out – we remember for example that we were going to take out the pastry from the freezer to defrost for the pie we planned to make for dinner. If we simply cannot remember what it is that we were doing we may decide to do something else instead – go and fold the clean washing say – or until it comes back to us. But then, in folding the washing, we once again comprehend what we are doing.

Still, I might realise *what* I am doing but not understand *why* I am doing it. I find myself ironing the sheets. But I never iron; doing so now just does not make sense. Velleman argues that “...people are generally guided in their behaviour by a cognitive motive towards self-understanding”.⁶ This motive stands behind reflection. It is not a desire for anything in particular – like the desire for a scotch and soda, or a warm bath, or to talk to my mother, or to take up salsa dancing. But it provides an important basis for us to discriminate between our various desires, since acting on some, but not other, of our desires will make more sense to us and so will appeal to our cognitive motives. What will make sense to us will depend on a range of variables: some will be situational/context dependent, and some more enduring. Our particular actions are undertaken in accordance with both short term and longer-term intentions, plans, and desires, and will make sense to us when framed as part of those plans. They can also make sense or not in terms of a person’s self-image.

So, I might want a warm bath now just as much as I want a scotch and soda but it would make more sense to have the scotch and soda given that my dinner guests are arriving in fifteen minutes and I still need to put out the hors d’oeuvres. A scotch and soda will fit with my schedule. A bath would not. But even setting aside the need to get the hors d’oeuvres out – suppose I have put them out already – it would be very odd, at least for me, to be in the bath when my guests arrive. I am not at all Bohemian or unconventional and I pride myself on being an excellent hostess. It simply would not make sense in terms of my self-image.

Velleman points out that a person’s self-image is to a significant extent self-fulfilling. Thinking of oneself as being a person of a particular kind, as shy or determined or hardworking say, or as having particular interests or skills, or as occupying particular roles, mother, teacher, researcher, or friend, can causally influence what we decide to do. It does this in part by influencing what it would *make sense* for us to do given who we believe we are. Our self-image grounds and delineates a set of considerations in the light of which we have reasons to do some things but not others. The more central a trait, or role, or value, or

⁴ Obviously, this is different to being unaware of what one is doing in the sense of being incompetent at the activity or not understanding the true nature and consequences of the action, for example, I think I am switching on the security system but I am actually disabling it.

⁵ Velleman’s example was forgetting what you are doing while ‘walking up Fifth Avenue’. See David Velleman, “Practical Reflection,” *The Philosophical Review*, XCIV, 1 (1985): 33–61, at p. 33. Psychologists call the phenomenon the *location updating effect* (or sometimes ‘doorway effect’). See Gabriel Radvansky and Jeffrey Zacks, “Event Boundaries in Memory and Cognition,” *Current Opinion in Behavioural Sciences*, XVII (2017): 133–140, at p. 136.

⁶ Velleman, *Self to Self*, *op. cit.*, p. 8.

activity is to our self-image or self-story the more powerful a factor it is in determining what is relevant and salient in reflection and which of the options available to us, we will take.⁷

What we *desire* to do in the moment and what it would make best *sense* for us to do in the light of our self-image can thus come apart and when they do our fundamental interest in sense-making may trump our particular desires.⁸ So, for example, my self-image as an honest person may make it inconceivable to me that I would cheat on my tax return, even though as it happens I would very much like some extra cash. It is not simply that I believe it would be wrong to do it – though I do – but that in conceiving of myself as rigorously honest, cheating becomes incomprehensible. Cheating would not cohere with my self-image. Or to take a rather different example: many people who have long-term addictions no longer gain pleasure from their substance use and may acknowledge that they should quit, given the negative consequences to their health and well-being. But using better fits with their stigmatised self-image. It is what they have learned to expect that they will do and so what they can understand themselves doing. Sobriety is like a foreign country where they would not know what to do or expect. It is a discomfiting prospect. As one person attempting to quit said:

I feel myself when I'm using and it's when I don't use... I don't feel myself. (...) But I'm trying to ... find myself without using, it's hard.⁹

Self-image then, is a powerful factor in sense-making and a powerful factor in diachronic agency since by its nature it persists whereas desires come and go. The more strongly internalised features of the self-image are, the more pressure there is on the agent to choose actions, plans and projects that are consistent with it in order that they remain comprehensible both to themselves and others, and in order to avoid the psychological discomfort of inconsistency and unpredictability. It is in this way that a stable self-image organises our agency over time, and it is no surprise then that people lacking a strong and stable self-image – including people with Borderline Personality Disorder, and people with psychopathy – display significant impairments of diachronic agency.¹⁰

⁷ Velleman outlines the details of the moral psychology of this view of agency in chapter 10 of *Self to Self*, *op. cit.*, pp. 224-253. In social psychology, the Social-Cognitive model identifies mechanisms linking the self-image with self-regulation, focused attention, the interpretation of situations, and response-dependencies, once incoherencies in motive and belief are eliminated. See Karl Aquino and Americus Reed, "The Self-importance of Moral Identity," *Journal of Personality and Social Psychology*, LXXXIII, 6 (2002): 1423–40.

⁸ Of course, both of these can come apart from what we judge best.

⁹ Doug McConnell and Anke Snoek, "The Importance of Self-Narration in Recovery from Addiction," *Philosophy, Psychiatry & Psychology*, XXV, 3 (2018): 31-44, at p.37.

¹⁰ We are not suggesting that these conditions are comparable in any other way. BPD is accompanied by significant personal distress, emotional intensity, and emotional dysregulation. Psychopathy is not. Nevertheless, both conditions illustrate the idea that the degrees of successful effective agency are expressed commensurately with the degrees of persistence and coherence in the self-image. Why is this true? Those with BPD or Psychopathy, and other agents, for example, those with dissociative psychopathologies, can be synchronically effective – say in attending an appointment, going on a day trip, or cooking a meal – but the completion of certain kinds of enduring projects requiring extended agency will be difficult, even out of reach; an unstable or changing self-image undermines the focus needed to develop the required skills and expertise. Moreover, some self-images – for example, those associated with professional roles such as scientist, lawyer, doctor, or teacher – require long-term coherence and stability. The effectiveness of what these agents do at *each* time depends on its contribution to an enduring end or project. The reason, roughly, is that at each time the diachronic agent must buy into the diachronic role (and accompanying self-image) just to be effective in that role. Of course, it might be claimed that there can be successful effective agency at each of the moments

II. SENSE-MAKING, TRUTH-MAKING, AND AGENCY

In the last section we explained how the drive for sense-making articulates with agency, via the self-image it supports. This is the self-fulfilling aspect to agency: in planning and projecting myself into the future, some subsets of reasons make more sense than others, given my self-image. Such a picture of ourselves, and of agency across time, is one of coherence between the self-image and one's desires, plans and intentions. However, successful cohesive agency depends on more than internal coherence; it depends, much of the time, on having both a true picture of the world and an accurate understanding of who we are within it. In this section we explain how sense-making links with agency, via a self-image and self-narrative that is anchored by an adequate understanding of facts that are relevant to it.

In general, we say that effective and successful diachronic agency – which involves a coherent comprehension of what one is doing and the reason for doing it – requires that an agent has a factually based self-understanding.¹¹ Contrast the case where I enter the library to borrow a book for my modest research project in my role as an academic, with the case where someone borrows the book while experiencing the psychotic delusion that they are a world-famous physicist, on the verge of a breakthrough that will allow us to make ourselves invisible. In each case, if we focus narrowly on the book-borrowing, the person knows *what* they are doing – borrowing a book – and can be successful in their *book-borrowing*. But though both have a story, which relies upon their self-understanding, to tell about the *why* – it is clear that only one of them can be successful in their diachronic project. If our self-understanding has no foundation in fact, our agency is diminished. We will not be able to do what we set out to do.

This factive requirement on self-understanding, as ordinarily required for successful agency, is similar to the requirement for truthfulness in our social interactions. Agency is relational and normative: we rely on input and feedback from others to make sense of the world and ourselves and in order to plan, to act, and to complete our projects. A world in which we are unable to rely on what others tell us as a basis for planning and action is unstable, and ultimately socially unsustainable, since it would undermine joint, as well as individual, projects. In the normal case, effective and successful agency is facilitated when a

in the life of an individual despite change over time in self-image. (We thank an anonymous referee for raising this point.) This possibility is compatible with a continuum of cases from the kind of agent who, over a lifetime, has a sequence of different self-images, each relatively enduring, to one whose self-images come and go in relatively rapid succession. This latter kind of agent is a possible agent, and the psychopathologies we have discussed may even describe such agents. For this kind of individual whether their agency can be successful and effective at each moment depends on the kind of project being undertaken. Some projects are integral to certain self-images – healing and being a doctor say – whereas some projects are generic – I may retain an interest in building my house despite changes in my self-image. An individual whose self-images change rapidly is, however, less likely to be an effective agent who attempts projects the completion of which is integral to long-lasting roles. It would also be difficult to complete long term projects when the interests and motivations of different sequenced (or partially overlapping) self-images are in tension.

¹¹ By 'self-understanding' we do not mean self-knowledge in a strict sense, since our self-understandings will almost certainly be in some respects based on false beliefs. We are employing a more ordinary and modest notion in which central beliefs I have about myself should be true. My name, age, address, family situation, job, roles, projects, education, likes and dislikes, central life events, and so forth all constitute facts about me that shape my agency. And being in touch with other everyday facts about the world also frames my self-image and my agency.

person's self-understanding is sufficiently accurate, and this accurate understanding is supported in our social interactions.

Our account of how sense-making links with agency via an accurately understood self-image is helpfully framed by outlining a three-way distinction in the normative requirement for truth as it relates to self-narratives. First, our stories should fit with how the world is – they should be *truth-tracking*. It ought to be the case that I did go to that conference I claimed to attend, that I did apply for the promotion that I am disappointed not to have got, that I do indeed have two children, and not five. Second, in our interactions with others, when we communicate what we have done, or we jointly plan what to do, we should do so truthfully: our stories ought to be *truth-shareable*. They should be such that others can rely on them and incorporate them into their own narratives. Third, it is, however, critical to note that the demand for our self-narratives to track what is (currently) true – its truth-tracking aspect – is modifiable by the extent to which our narratives can be *truth-making*.¹² Let us return to the case of addiction. For many people with addiction, recovery involves a deliberate and active change but the newly adopted self-image may not be one that is yet supported by the evidence – unlike their addict identity, it is not truth-tracking. Yet it can be truth-making. Velleman recounts the story of a colleague who successfully gave up smoking by ceasing to think of himself as a smoker: "...he then enacted what he was imagining, pretending to be the non-smoker that he wanted to be".¹³ As a non-smoker buying and smoking cigarettes no longer makes sense – it is clearly not something that a non-smoker has any reason to do – and so desires to smoke and the discomfort of withdrawal could no longer be foregrounded in deliberation as they would be for the smoker trying to quit. Such desires and feelings would appear as irritations rather than as temptations. Velleman suggests that: "The point of identifying with...the non-smoker was precisely to gain access to a different story, presenting a different set of reasons."¹⁴ The story of oneself as a non-smoker, or as kind, or honest, or outgoing, or health-conscious, can be made true, because of the role of self-image in agency via our concern for sense-making.

Of course, there are limits to the ways and the degree to which we can change our self-image. When we regret the type of person we have become, striving to be a different kind of person, our view of what we have most reason to do may diverge from what we can most easily understand ourselves doing.¹⁵ In the case of addiction, recovery-directed plans can feel alienating and unbelievable because they do not cohere strongly enough with one's history. The drive for sense-making can thus dispose one to relapse. As McConnell and Snoek say: "the less plausible it is that a new narrative thread can be a continuation of the established self-narrative, the greater the feelings of self-alienation".¹⁶ It can take time, and

¹² See Steve Matthews and Jeanette Kennett, "Truth, Lies, and the Narrative Self," *American Philosophical Quarterly*, XLIX, 4 (2012): 301–15, at p. 306.

¹³ Velleman, *Self to Self*, *op. cit.*, p.325. Several studies support this notion. See, for example, Ildiko Tombor, Lion Shahab, Jamie Brown, Caitlin Notley, and Robert West, "Does the Non-Smoker Identity following Quitting Predict long-term Abstinence? Evidence from a Population Survey in England," *Addictive Behaviors*, XLV (2015): 99–103, at p.100, where they remark that "[t]his study assessed the prospective associations between taking on a non-smoker identity following quitting and long-term abstinence...Ex-smokers who make this mental transition following a quit attempt appear more likely to remain abstinent in the medium term than those who still think of themselves as smokers." See also Eline Meijer, Winifred A. Gebhardt, Collette van Laar, Bas van den Putte, and Andrea W. M. Evers, "Strengthening Quitter Self-identity: an Experimental Study," *Psychology and Health*, XVIII, 10 (2018): 1229–50.

¹⁴ Velleman, *Self to Self*, *op. cit.*, p.327.

¹⁵ We thank an anonymous referee for raising this difficulty.

¹⁶ McConnell and Snoek, "The Importance of Self-Narration in Recovery from Addiction," *op. cit.*, p.35.

what McConnell and Snoek call “narrative work”, to overcome this gap in coherence. This will involve identifying narrative threads in one’s history that can be woven into new and more hopeful stories about oneself that support and cohere with recovery plans.¹⁷

With this proviso in place, we now describe some ordinary ways in which sense-making can be either supported or undermined in our relations with others and consider how this interacts with adherence to the truth.

III. FACILITATING AND BLOCKING SENSE-MAKING: FRIENDSHIP AND GASLIGHTING

For most of us our sense of who we are and how we are faring is dependent to some significant extent on social feedback. Our self-image and self-narratives are not constructed in splendid isolation. Other people – particularly close others – support us in our sense-making project.

To see this, consider the case of close friendship. Mutual interpretation is, plausibly, a constitutive feature of intimate relationships such as friendship.¹⁸ This is both a self-making and sense-making activity. In the reciprocal sharing, narrating, interpreting and affirming of events, traits, quirks, activities, and potentials, close friends provide us with a picture of ourselves that confirms or can be taken up into the self-image to explain and help organise our doings. Friends actively support each other in sense-making. In doing so we say they manifest recognition and respect for each other’s agency. It is important to note, however, that in order to support sense-making, mutual interpretation should be anchored in truth. It must be truth-tracking or truth-making. I do not do my friend any favours, nor do I show her respect, if I encourage or endorse a totally unrealistic view of events and one moreover that I do not believe myself. But, as in the case of changing an addiction narrative, I might through my interpretations encourage her to be bolder, more playful, more generous, or more resolute. I might for example highlight some aspects of her history or character and downplay others to help her to see that a significant career change or the ending of a relationship can be a meaningful continuation of her story rather than an abrupt rupture of it.

But just as we can facilitate each other’s sense-making, so, in myriad ways, we can block or undermine it. Our claim is that conscious or reckless blocking or undermining of another’s sense-making efforts is a clear failure of the respect owed to them as agents. The most vivid illustration of this is found in the egregious phenomenon of gaslighting that occurs within some intimate relationships and also often in broader professional and social situations.

Kate Abramson characterises gaslighting this way:

Very roughly, the phenomenon that’s come to be picked out with that term is a form of emotional manipulation in which the gaslighter tries (consciously or not) to induce in someone the sense that her reactions, perceptions, memories and/or beliefs are not just mistaken, but utterly without grounds—paradigmatically, so unfounded as to qualify as crazy. Gaslighting is, even at this level, quite unlike merely dismissing

¹⁷ See also Doug McConnell, “Narrative Self-constitution and Recovery from Addiction,” *American Philosophical Quarterly*, LIII, 3 (2016): 307–322; Hanna Pickard, “Addiction and the Self,” *Nous*, LV, 4 (2021): 737–761; Jeanette Kennett, “Just say No? Addiction and the Elements of Self-control,” in Neil Levy, ed., *Addiction and Self-control* (New York: Oxford University Press, 2013), pp. 144–64.

¹⁸ See Dean Cocking and Jeanette Kennett, “Friendship and the Self,” *Ethics*, CVIII, 3 (1998): 502–527.

someone, for dismissal simply fails to take another seriously as an interlocutor, whereas gaslighting is aimed at getting another not to take herself seriously as an interlocutor.¹⁹

Typical phrases used by the gaslighter include, “You’re imagining it!” “You’re paranoid”, “Don’t be so sensitive”, “You’re overreacting”, “I was joking”, and so forth.²⁰ The person who is the target of gaslighting loses confidence in themselves: in their perceptions and their judgments, their skills and their personal traits. They are portrayed to themselves as paranoid, jealous, oversensitive, maybe even crazy. They fear they are losing their mind. This attack on their self-image profoundly affects their capacity to make sense of the world and of themselves. It affects their capacities for planning and action since they are no longer sure about who they are.

Abramson gives the example, from a film, of Pat, a skilled and ambitious golfer, whose confidence is systematically undermined by her husband Collier, who wants her to be a stay-at-home wife. She says:

Notably...the sense in which Pat ends up undermined isn’t just about her golfing abilities – she says she feels “carved up”, “nobody”. This kind of language is common among targets of successful gaslighting. It’s in the same category, for instance, as De Beauvoir’s [response to Sartre’s criticisms] “I am no longer sure...*even if I think at all*”. *It’s language that speaks to a sense of having lost one’s independent standing as deliberator and moral agent* (our italics, last sentence).²¹

Gaslighting can also take place at the institutional and societal level. For example, the Kafkaesque rules imposed on social welfare recipients can lead them to doubt their competence and worth; their representation in the media and by politicians as ‘welfare cheats’ and ‘work shy’ systematically devalues their lives and discounts their perspectives. Like Pat they lose control of the meaning of their actions, and their moral status and agency is thus diminished. Gaslighting is a profound failure of respect that involves a dismissal of the dignity and moral status of its victims.²²

Of particular note here is that successful gaslighting at both the personal and institutional levels need not rely on lies or deception. A failure to be truthful is not the fundamental source of the disrespect shown in such cases. An impossible bureaucratic maze can disrupt and undermine agency without deception. Given a background of patriarchy, a man like Collier might comfortably believe that his wife’s golfing prowess is a flash in the

¹⁹ Kate Abramson, “Turning Up the Lights on Gaslighting,” *Philosophical Perspectives*, XXVIII, 1 (2014): 1-30, at p. 2.

²⁰ *Ibid.*, p.1.

²¹ *Ibid.*, p. 8.

²² In similar vein Jeanette Kennett and Jessica Wolfendale discuss the impacts on agency of the loss of, or absence of moral security. See Jessica Wolfendale and Jeanette Kennett, “Self-control and Moral Security,” in David Shoemaker (ed.), *Oxford Studies in Agency and Responsibility Volume 6* (New York: Oxford University Press, 2019), pp. 33–63. They note (p.42): “[w]e can injure each other by failing to give each other due recognition, respect, and goodwill (as well as by intentional malicious acts). By their actions and their attitudes, others can vividly demonstrate to us that they do not see us, or the group with which our identity is bound up, as morally significant; that our pain and our suffering are not important, or that our values and choices are morally inferior.” See also Jessica Wolfendale, “Moral Security,” *The Journal of Political Philosophy*, XXV, 2 (2017): 238–255.

pan that will get her nowhere, inappropriate in any case for a woman, and that she will be happier as a stay-at-home wife. He knows better than her and so her perspective may not only be dismissed by *him*. It *should* be dismissed by her. It is of little account.²³ The charges levelled by the gaslighter, of overreaction, oversensitivity, or lack of a sense of humour, might be sincerely felt. Nevertheless, they fail to display respect and recognition of the other in ways which directly undermine sense-making.

Gaslighting is an extreme example of what can happen when our sense-making capacities are attacked. But it serves to highlight the vulnerability of the self-image and the severe impacts on agency when it is undermined. It also further highlights the important role friends and close others usually play in supporting and scaffolding sense-making, and which we argue is a requirement of respect.

IV. TRUTHFULNESS, AND THE COSTS TO AGENCY OF NON-TRUTHFULNESS

To recap, we have now set out an important norm arising from within a relational standpoint. In our relations with others – friends, partners, or professional relations – respect for agency will be partly constituted by an active concern to understand the perspective of the other in order to see how the world makes sense to them and to support their capacity to do so within the relevant sphere. And we have suggested that ordinarily, truthfulness is a condition of successful sense-making. In this section we set out in more detail the norm of truthfulness: what is at stake morally in adhering to it, and why we ought, other things being equal, to maintain it. In the section that follows this one we address cases where respect for sense-making and respect for truthfulness appear to be in competition.

In our relations with others, the norm of truthfulness, with its central demand for honesty, is fundamental. For example, in the professional sphere, respect for persons and their agency requires truthful sharing of information, because often enough it forms the basis upon which to make rational decisions. Truthfulness is a prerequisite for client autonomy and informed consent. Even if we do not fully subscribe to the Kantian view that lying *always* fails to respect persons as ends-in-themselves, we nevertheless show disrespect for another when rationing their access to information, or substituting our own judgment for a decision that ought to be theirs. Truthfulness matters in both personal and professional settings for informed planning and decision-making and is thus critical to effective diachronic agency.

Being reliably truthful is also essential in close friendship and intimacy. Self-understanding and the mutual interpretation process described above require truthfulness. To see the importance of this we only need consider cases where others are not truthful. Being lied to, especially in systematic ways, can undermine our moral position in the world.

To bring this out consider the case of Jean-Claude Romand.²⁴ Romand began systematically deceiving those around him – including his immediate family, his parents, and

²³ Note that such attitudes which talk down are also a way of shutting down or limiting their target's possibilities for truth-making. "Women don't golf"; "If you do that people will think you are ..." thanks to Doug McConnell for this point.

²⁴ See Emmanuel Carrere, *L'Adversaire* (New York: Picador, 2000). We discuss the case at length in "Truth, Lies, and the Narrative Self," *op. cit.*, p. 302. There are other cases of spectacular deceptions, for example, Donald Crowhurst (https://en.wikipedia.org/wiki/Donald_Crowhurst), whose failure to complete an ocean yacht race led to his faking a circumnavigation. Cases can be found also of people who claim to have fought in wars when

some friends – after claiming to have passed an exam for medical school which he failed to attend. He continued the deception in spectacular ways, claiming that he received his degree, and that he was working as a medical professional and researcher at the World Health Organization. He would claim to go on work trips, but in fact he would attend the local WHO building or Geneva International Airport. There he would study travel brochures and medical journals so as to keep up the façade. This life carried on for nearly twenty years. At the end, fearing that the truth was about to be revealed Romand killed his family and parents.

The theoretical interest in such cases is twofold. First, failures of truthfulness affect the possibility of intimacy. Romand could not disclose his true self to others, and this blocks the goods he (and those around him) derive from close relationships. A normal mark of intimacy is the truth-shareability of our self-narratives and this depends on our freely, openly, and tractably sharing our stories. In cases of living a lie, in which people systematically misrepresent themselves to others, the norms of truthfulness mentioned above – where our stories are shared as being truth-tracking or truth-making – fail to be observed and this presents a barrier to intimacy. The problem is that the narratives of those who live a lie are “frozen”. Each telling of the story of their false past involves the need to be mindful of consistency lest they be caught out. Normally when we talk about our pasts truthfully, we do so openly and with a preparedness to be corrected, and a willingness to re-interpret our remembered experiences. Romand, and others like him, are unable to relate to close others with such openness. So, Romand’s self-understanding cannot be supported via a process in which others can get to know him better through exchanges of personal information that can be talked about, reflected upon, discussed, revised, celebrated, lamented, and so on. This restricts a person’s capacity to jointly and flexibly make sense of themselves and others, something that is normally required for the healthy development of a person both for their own sake and the sake of their close relationships. And given this constriction around their capacity for self-development, there is a corresponding diminishment of their agency.

Second, deception impinges on the capacity of the *other’s* self-image to be grounded in the truth. Romand’s wife presumably built her self-image and activities in significant part on being the wife of a medical researcher who travelled a lot. Had she discovered the truth she may well have felt lost, unanchored, and incapable of making sense of the past twenty years of her life. As in the much more common case of Chris and Jo, with which we began, if Jo discovers Chris’s long running infidelities, she will question all her past assumptions and interpretations and may wonder: Who am I really? Was anything as it seemed?

V. TRUTHFULNESS VERSUS SENSE-MAKING IN SUPPORTING AGENCY

Sometimes we may permissibly diverge from being truthful. The standard justification for such divergences is typically to avoid some significant harm. So, with no evidence, I tell my very drunk friend flaunting with the idea of taking a night-time dip in the sea that there is a shark in the water. I know this will deter him, where other prudential considerations will not. Lies may at least sometimes be permissible to avoid harms, but telling such lies is nevertheless a vehicle designed to manage or manipulate others and as such falls short of respect for their agency. In the case of the drunk friend the offence is minor; the overriding

they did no such thing, for example, the case of Rex Crane. See Simon Caterson, *Hoax Nation: Australian fakes and frauds, from Plato to Norma Khouri* (Melbourne: Arcade, 2009).

demand is their protection at a moment when their agency is impaired. It is, presumably, a one-off. Though it manipulates the friend, it is motivated by friendship. The main moral danger might be that manipulating the friend becomes a habit, and is then genuinely or seriously disrespectful of the agency of that person.

Nevertheless, are there cases where *respecting* another's agency requires or permits us to set aside the normative demands of truth? Here we are thinking of cases where a person might make sense of their environment and act in ways that cohere with their self-understanding thus delivering the associated benefits to agency, but the premises on which they construct or maintain their self-image fail to be truth-tracking and cannot be truth-making. If so, what does respect for agency demand from those around them? And what are the implications for intimate relationships?

Our position is that while truthfulness should be our default stance, whole truth-telling, or strict insistence on what is true should not always trump practices that support the other in sense-making. To explore the way these two modes of respect for agency apply we present four case-types:

V.1. The Two Friends. Consider two friends, one of whom is an atheist and the other deeply religious. Each must believe the other to be deeply mistaken in respect of religion. The religious person's beliefs are core to her self-image and self-narrative and constitute an organising principle in her life. As well as religious observances, she adheres to dietary considerations, particular familial obligations and so forth. Now while the friends may well openly discuss their different beliefs and the reasons for them, it would seem like a clear failure of respect were the atheist to insist that her religious friend's beliefs are no better grounded than a fairy-tale, or to try to dissuade her from going to church and so forth when these things play so central a role in her life. Similarly, it would be disrespectful were the religious friend to insist that her friend's atheism rendered her incapable of moral concerns and to remind her friend on a daily basis that she is going to hell unless she converts. And because friends are also directed by each other's concerns and interests the non-religious friend will likely go further than merely not interfering. She might well, out of friendship, support in various ways her friend's religious pursuits, say by driving her to church if her friend's car has broken down, or remembering to bring her special food for religious feast days. Though she thinks her friend's way of life is based on beliefs that are strictly false, her love for her friend and her recognition of the centrality of her friend's religious commitments to her life is more important than a rigid insistence on what she believes to be true. She might even conclude that, were it possible for her to convince her friend to give up her faith, she ought not to do so. Were her friend's religious faith to be upset it would most likely cause severe disruption to her agency. The religious friend may quite literally not know what to do without the framework provided by her self-image as a believer.

Perhaps, in such cases, we might say that what is being accepted is her friend's self-conception. She is a religious person, and it is this fact about her that is being respected along with the organising role it plays in her agency. We think that such allowances are not uncommon between friends and family members when it is judged that there are other important goods at stake, and that they are seen as a requirement of respect rather than a violation of respect.

V.2. Joint Remembering. There are also other utterly familiar kinds of cases where a rigid insistence on accuracy can interfere with overall truth-tracking and impair the potential for both truthfulness and sense-making. Consider first the case of joint or collaborative remembering.

Remembering is often something we do together, and memory research has shown that remembering is not a strict truth-tracking exercise. Rather, as Schectman has argued, memories are constructed through a process that involves condensing and summarizing information about the past into representative events, remembering some events in detail, some vaguely or not at all, contextualizing, threading events into narratives, and reinterpretation of the meaning of past events.²⁵ Our subsequent recounting of the past may not tell the whole truth, and will likely also contain inaccuracies, yet succeed in capturing the gist and importance of events or a series of events, conversations, relationships, moods and perspectives. Memory thus has a critical sense-making role – it creates usable self-knowledge and a springboard to the future. We pick and choose what we take to be the most important elements and key moments of the past in explaining and expressing who we are.

As Harris et al note, joint or group remembering “...is an interactive activity where memories are dynamically and jointly constructed in conversation ... These conversations are one way that groups develop shared memories of the past.”²⁶ We add details and context to each other’s stories. “It was on a Wednesday”. “Was Bill there too?” “Yes! We ate oysters.” “It was a beautiful day, but Pearl seemed distracted”. “Oh, that must have been because, remember we found out later, she had just lost her job.” This kind of cooperative exchange that scaffolds and enriches remembering can be derailed by an over-zealous insistence on accuracy by one of the conversational partners. Harris et al have found in their research with older couples that constant corrections derail effective collaboration leading to a reduction in remembered content. Insistence on truth-tracking can thus damage truth-sharing. Independently of the implications for remembering, however, they suggest that social and relational benefits are lost when a conversational partner is intolerant of perceived inaccuracies in the other’s story. First and most obviously it has the predictable effect of shutting down interaction. Presumably the partner who is interrupted may also feel resentful. Now it might be said in such cases that the normative demand for truth is simply in conflict with social and utilitarian considerations. You are sure that the restaurant dinner for June’s birthday was not the time Bill made a scene but it is not worth upsetting the relationship by continuing to insist on it. It might thus be argued that these kinds of cases should be assimilated to the case where I lie to my friend in order to prevent him from swimming while drunk. We think this would be a mistake. *Constant* interruption and corrections evince a lack of respect for the other. It is similar to gaslighting in that it fails to take the other seriously as an agent and tends to undermine their self-confidence. The source of the damage to the relationship lies in the disrespect which blocks, rather than facilitates, the person’s attempts to reconstruct and make sense of the past.

V.3. Confabulation. Constant correction and insistence on rigorous truth-tracking can disrupt, rather than assist, the reconstructive and sense-making work of memory in part because of the ubiquity of forms of *confabulation*. There are a variety of theories to explain confabulation, with some focusing on the neuropsychological deficits, and some embedding confabulation within a theory of delusions.²⁷ In this paper we focus on confabulation in so far as it plays a kind of epistemic role, where the fictionalising response fits with the

²⁵ Marya Schechtman, “The Truth about Memory,” *Philosophical Psychology*, VII, 1 (1994): 3-18.

²⁶ Celia B Harris, Paul G. Keil, John Sutton, Amanda J. Barnier, and Doris J. F. McIlwain, “We Remember, We Forget: Collaborative Remembering in Older Couples,” *Discourse Processes*, XLVIII, 4 (2011): 267–303 at p. 268.

²⁷ See Michael D. Kopelman, “Varieties of Confabulation and Delusion,” *Cognitive Neuropsychiatry*, XV, 1 (2010): 14–37.

normative social demand for sense-making, and is also a response to the fundamental agential need for intelligibility of the sensed world. Interpreted this way, confabulations that are observed after brain damage fulfill the function of establishing a psychological bridge between what is remembered and what is not, in a way that is often seen as an attempt to preserve a positive self-image. As Bortolotti et al nicely state the point:

[Confabulation]... allows people to keep constructing self-narratives in situations where personal information is not available. As a result, it secures some psychological continuity with the confabulators' previous selves in the absence of reliable recollective capacities, and contributes to the preservation of psychological integration in absence of introspective access to the reasons for conscious mental states such as beliefs, desires and preferences. It also allows people to include new facts into previously developed narratives that have become fragmented.²⁸

There is a continuum between the phenomenon associated with brain damage through to everyday cases of narrative retelling where confabulation occurs to plug gaps in autobiographical memory. The functions of confabulation can also be observed in normal subjects who tend to present their current selves in a way that is both coherent and largely favourable.²⁹ We commonly and spontaneously construct fictive bridge sections to cover gaps in autobiographical memory, or we enact "scripts" that fit with our current self-image. Our sense-making motives drive this everyday confabulation. In a discussion of the literature on confabulation Spear notes that

Confabulators don't tell just any story, they tell a story that closes a gap in their beliefs, and they typically do so in a way that maintains consistency among their beliefs and in their conceptions of themselves... maintaining such coherence of self-concept in the case of confabulation typically comes at the expense of truth, challenging the idea that it is epistemically beneficial. However, ... [it] may sustain or support epistemic benefits *indirectly*.³⁰

What are these benefits? Autobiographical understanding is typically intolerant of gaps, inconsistencies or tensions more generally, say in one's principles, attitudes or emotions. Confabulations appear to address this intolerance; they may even function as an aid in the elimination of distortions in one's sense of self, or self-image. As Gallagher says: "To some degree, and for the sake of creating a coherency to life, it is normal to confabulate and to

²⁸ Lisa Bortolotti, Rochelle Cox, Matthew Broome, Matteo Mameli, "Rationality and Self-knowledge in Delusion and Confabulation: Implications for Autonomy and Self-governance," in Lubomira Radoilska, ed., *Autonomy and Mental Disorder* (Oxford: Oxford University Press, 2012), pp. 100–23.

²⁹ See Anne Wilson and Michael Ross, "The Identity Function of Autobiographical Memory: Time is on our Side," *Memory*, XI, 2 (2003): 137–49.

³⁰ See Andrew D. Spear, "Gaslighting, Confabulation, and Epistemic Innocence," *Topoi*, XXXIX, 1 (2020): 229–41, at p. 555, and p. 557. Spear's very interesting paper focuses on cases, specifically gaslighting, where he thinks confabulation often occurs. Confabulation is not always epistemically innocent as the literature on implicit bias demonstrates. (See, for example, Ema Sullivan-Bissett, "Implicit Bias, Confabulation, and Epistemic Innocence," *Consciousness and Cognition*, XXXIII, (2015): 548–60). Some of the desires or interests that motivate confabulation, for example, of the reasons why a male candidate was more appointable than the identical or better qualified female candidate, are truth-undermining and are so in ways we think that are criticisable.

enhance one's story."³¹ Ramachandran suggests that the function of confabulation, "...is to create a coherent belief system in order to impose stability in one's behaviour...When something doesn't quite fit the script...you very rarely tear up the entire story and start from scratch. What you do, instead, is to deny or confabulate in order to make the information fit the big picture."³²

And as Bortolotti and Cox have it, "...confabulation may allow subjects to exercise some control over their own cognitive life which is instrumental to the construction or preservation of their sense of self."³³ In other words, they can be an aid to sense-making, and in that respect, far from being harmful they can be enabling of agency. Bortolotti develops this in terms of the way confabulations can be *epistemically innocent*, and proposes the following two conditions:

1. Epistemic Benefit: The delusional belief confers a significant epistemic benefit to an agent at the time of its adoption.
2. No Alternatives: Other beliefs that would confer the same benefit are not available to that agent at that time.³⁴

The epistemic innocence of fictive bridges between remembered events would obtain provided that they are truth-facilitating or at least what we term *truth-neutral*. That is, they do not affect the *overall* truth of the narrative; they merely allow the narrative to proceed by connecting true events in a way that makes sense.

A strict demand for truthfulness in response to such ordinary and ubiquitous incidents of confabulation would thus miss the moral mark by a considerable distance, and that is partly because those who confabulate are not intending to deceive their audience (this is a definitional requirement of confabulation), and partly because confabulation is functioning here as a way of meeting the demands of sense-making, at least for the person who engages in it. That there are such enabling sense-making fictive plugs in everyday narratives shows that the normative demand for truthfulness is not unlimited and, in some circumstances, a rigid insistence on truthfulness will be a mark of disrespect for the other and detrimental to their agency.

V.4. Dementia. Respecting agency through support for sense-making, even at some cost to truthfulness, can be seen to reach its ultimate limit in certain cases of dementia. There is a subset of cases in relatively advanced Alzheimer's Dementia in which people's self-images persist despite severe deficits in episodic memory function.³⁵ In these cases, people are marooned in the past and they attempt to build a bridge with their premorbid

³¹ Shaun Gallagher, "Self-Narratives in Schizophrenia," in T. Kircher and A. David, eds., *The Self in Neuroscience and Psychiatry* (Cambridge: Cambridge University Press, 2003), pp. 336-57, at p. 348.

³² V. S. Ramachandran, "The Evolutionary Biology of Self-Deception, Laughter, Dreaming and Depression: Some Clues from Anosognosia," *Medical Hypotheses*, XLVII, 5 (1996): 347-62.

³³ Lisa Bortolotti and Rochelle E. Cox, "'Faultless' Ignorance: Strengths and Limitations of Epistemic Definitions of Confabulation," *Consciousness and Cognition*, XVIII, 4 (2009): 952-965, at p. 952.

³⁴ Lisa Bortolotti, "The Epistemic Innocence of Motivated Delusions," *Consciousness and Cognition*, XXXIII, (2015): 490-99, at p. 496.

³⁵ See Stanley B. Klein, Leda Cosmides, and Kristi A. Constabile, "Preserved knowledge of Self in a Case of Alzheimer's Dementia," *Social Cognition*, XXI, 2 (2003): 157-65; D. C. Mograbi, R. G. Brown, and R. G. Morris, "Anosognosia in Alzheimer's Disease - the Petrified Self," *Consciousness and Cognition*, XVIII, 18 (2009): 989-1003; Steve Matthews, "Moral Self-Orientation in Alzheimer's Dementia," *Kennedy Institute of Ethics Journal*, XXX, 2 (2020): 141-66.

self; they construct a story that attempts to make sense of their present surroundings, experiences, and feelings in the light of their retained self-image (largely based around a role-conception). Their carers face the (sometimes) difficult choice of allowing, or even facilitating such bridge-building for the sake of agency, or alternatively to insist on being truthful and derailing such agency.³⁶ We think that in many of these cases such insistence would be fundamentally disrespectful, not just distressing for those persons.

Consider two actual cases:

Martha. Martha has Alzheimer's dementia and is attending the dayroom with several residents in a dementia care facility. It is the afternoon, and coffee has been served by the staff, who now have returned to the kitchen, but,

Somewhere along the way [Martha] assumes the status of gracious hostess – a part that has been hers innumerable times throughout her life. Equally innumerable are her stories about how she, like her mother before her, has taken pride in welcoming everyone to her home and in sharing her food and hospitality with other people...These narrative plots occur frequently in the body of material that can be described as variants of one overall storyline that binds together many of Martha's stories. This storyline is of her being a generous and sharing person throughout her life...³⁷

Thus, her role-driven self-image – one specifically infused with a moral identity of someone who is 'gracious', 'hospitable', 'generous', 'sharing' – emerges to fit with Martha's current experiences; she is able to make sense of the situation in a way that is continuous with her retained self-image and the narrative plots still available to her. To intervene here by correcting her poses a risk to her sense of self and her agency. Its effect no doubt would be to confuse Martha and bring what she took herself to be doing to an abrupt halt. In the circumstances, telling Martha the truth does not seem like a requirement, or mark, of respect for her.

Mr. Q. Oliver Sacks describes how a former janitor in a boarding school, Mr. Q, had developed Alzheimer's Disease and was residing in a care home run by the Little Sisters of the Poor.³⁸ Over time Mr. Q began to enact the janitorial role, checking that windows and doors were locked, inspecting laundry and boiler rooms to see that all was functioning well, and so on. As Sacks puts it,

The sisters...though perceiving his confusion and delusion, respected and even reinforced [his] identity...they assisted him in his janitorial role, giving him keys to certain closets and encouraging him to lock up at night before he retired...And, though he slowly became more and more demented over the years, he seemed to be organized and held together in a remarkable way by his role...Should we have told Mr. Q. that he was no longer a janitor but a declining and demented patient in a

³⁶ See Philippa Byers, Steve Matthews, Jeanette Kennett, eds., "Special issue: Truthfulness and Deception in Dementia Care," *Bioethics* XXXV, 9 (2021): 839–969.

³⁷ Linda Orulv and Hyden Lars-Christer, "Confabulation: Sense-making, Self-making and World-making in Dementia," *Discourse Studies*, VIII, 5 (2006): 647–73, at pp. 655–56.

³⁸ Oliver Sacks, "How Much a Dementia Patient Needs to Know: Should a Doctor Replace an Accustomed Identity with a Meaningless "Reality"?", *The New Yorker Magazine*, (2019). [How Much a Dementia Patient Needs to Know | The New Yorker](#).

nursing home? Should we have taken away his accustomed and well-rehearsed identity and replaced it with a “reality” that, though real to us, would have been meaningless to him? It seemed not only pointless but cruel to do so—and might well have hastened his decline.³⁹

In this case, as Sacks points out, insistence on the truth would be undermining of agency in the most fundamental way: it would remove Mr. Q’s capacity to make sense of his situation and to act in a coherent way on the basis of his role-understanding and beliefs about the kind of place in which he finds himself. Mr. Q was no longer capable of truth-tracking. But providing him with the tools and support for the janitorial role triggered narrative connections to his past and activated his retained self-image. As Strickwerda-Brown et al put it,

...the self is not lost in [Alzheimer’s Dementia] per se. Rather, the individual reverts to an older iteration of the self that is incongruent with their present experience and surroundings...one’s self-schema may operate as a dynamic, heuristically-driven template that facilitates fast decisions regarding “who I am”, “what I do”, and “how I behave” – rules that govern our day-to-day activities.⁴⁰

We suggest that in respecting the only self that remains, sense-making has clear priority over truthfulness in such cases. The sisters facilitated Mr. Q’s social agency by engaging with, rather than challenging, his out-of-date role identity. Without this implicit awareness of, and respect for, the importance of sense-making to agency and identity, Mr. Q may have ceased to act at all.

VI. TWO OBJECTIONS

We have argued that self-understanding is essential for effective human agency. We began with Velleman’s example of forgetting *what* one is doing. But a person may also forget *who* they are, and so may experience an absence of self. I wake in a strange place, I do not know where I am, and I do not, for a brief moment, have any sense of who I am. It is a paralysing and unnerving experience. Gradually it comes back to me, and I am able to make sense of my situation. Such experiences of total disorientation are rare but in Alzheimer’s Dementia they become more common and in the late stages of the disease a person may have no capacity to filter or organise their experiences and no, or very little, understanding of who they are (though they may retain some personality traits, preferences and procedural memory). These are not the cases we focus on. Mr. Q and Martha have not reached this point; they clearly retain some sense of who they are and are able to act in the light of this. But do they count as having self-understanding?

It might be objected, first, that neither Mr. Q nor Martha possesses self-understanding as we have described it, since relevant central beliefs they hold about themselves are false. At best they have a purported self-understanding. Since their self-understanding is significantly inaccurate it may follow that their apparent sense-making

³⁹ *Ibid.*

⁴⁰ Cherie Strickwerda-Brown, Matthew D. Grill, Jessica Andrews-Hanna, and Muirann Irish, “All is Not Lost – Rethinking the Nature of Memory and the Self in Dementia,” *Aging Research Reviews*, LIV, (2019): 1–11, at p. 4.

activities are likewise merely purported. This leads to the second objection. In cases of marginal or diminished agency it might be claimed that it is beneficence that does and should govern any decisions about whether to “go along with” the outdated self-image, rather than respect for the person’s sense-making needs and capacities. We address these objections in turn.⁴¹

VI.1. First Objection: Do Marginal Agents have genuine self-understanding? We do not require in the ordinary case that our self-understanding be fully accurate and in accordance with all the facts, but we have argued that truthfulness is the default normative standard. The exceptions to this requirement that we have surveyed are cases in which the person’s self-understanding is, for the most part, sufficiently well aligned with their circumstances, and where insistence on truthfulness would disrupt sense-making and be significantly disrespectful. However, cases of marginal agency undoubtedly test the boundary of our claim, and some philosophers will reasonably refuse to step across this boundary. For people like Mr. Q – let us describe marginal cases of this type as *remnant identity* cases – how do we characterize a self-understanding grounded in a true past, but one which is not able to be updated to match the person’s current circumstances? Mr. Q does not realise that his home is not a school and that despite his attention to janitorial tasks he is not really the janitor. The objector claims that he no longer possesses genuine self-understanding, given that he can no longer reliably engage in truth-tracking, and that what he does is thus merely purported sense-making.

We argue that the fact that Mr. Q’s beliefs about himself are *historically* true matters to the case. They have simply become unmoored from a point in time. It remains true that Mr. Q worked as a janitor, and it remains true that he enjoys and is (to some extent) capable of janitorial tasks.⁴² He likely, accurately, sees himself as someone who likes to be busy, useful, and so forth. A person living with dementia may retain an identity that is substantially historically accurate, and also one that accurately represents their values, tastes, and cares. In the case of Martha these character traits manifest in her gracious hostess identity. To be sure, aspects of the self will fade or be extinguished as the dementia progresses. The person’s capacity to *articulate* their self-understanding will degrade, but as dementia research indicates, the self is not altogether lost.⁴³ Moreover, their remnant identity continues to furnish them with a cognitive motive. The drive for sense-making is preserved.

For such marginal cases we suspect it is the lack of illness-insight that supports the case for thinking they have a merely purported self-understanding. The remnant identity

⁴¹ We thank two referees for pressing us on each of these objections.

⁴² Role identities persist beyond the period in which one occupies a role and in the case of many roles it is not always clear cut that the person can no longer be referred to by the role-name. Consider the retired nurse who has given up her professional registration who comes upon the scene of an accident or medical emergency. Is she wrong to say as she takes charge of the situation “I’m a nurse”? Or consider the retired professional philosopher who continues with some of their philosophical activities, someone who attends talks, engages with former colleagues, co-writes the occasional paper. There are real cases of persisting role-identities in dementia as well. Steven Sabat discusses the case of Dr. B, a retired scientist with mid to late-stage Alzheimer’s Disease, who met with him each week to work on a research project. Sabat remarked that ‘...in terms of his disposition, habits of mind, and his sense of who he was, he was very much a scientist...his proclivities were clear in the way he comported himself at the day center, so it would be utterly wrong to say about him “he used to be a scientist.”’ See Steven Sabat, *Alzheimer’s Disease and Dementia: What Everyone Needs to Know*, (New York: Oxford University Press, 2018), p. 70.

⁴³ See Strickwerda-Brown et al, “All is Not Lost,” *op. cit.*, p. 4.

cases we see in Alzheimer's fall on a spectrum of self-understanding, with insight progressively lost over the course of the disease. Positioning within the spectrum depends – in addition to facts about knowing who and where you are – on facts concerning such insight.⁴⁴ Mr. Q and Martha's lack of insight cannot be repaired because relevant neurological capacities are damaged. We concede that their remnant self-understanding is thus not *robust* self-understanding, or as we glossed it above (fn.11); but nor is it merely purported self-understanding. Mr. Q and Martha possess genuine, albeit simplified and historically fixed self-images.⁴⁵ They retain sufficient accuracy of their self-image to ground supportive practices that scaffold their agency, as well as some of the truth-making elements of mutual interpretation we discussed above in Section III.⁴⁶ We suggest in the light of this, that changes to self-understanding in our cases are analogous to the changes to valuing in dementia that Jaworska describes.⁴⁷

Jaworska rejects the view that valuing requires a person to have a sense of their life as a whole; in Alzheimer's she claims that 'valuing becomes uncoupled from the person's narrative of their whole life'.⁴⁸ Earlier values may be reconfigured to match fading cognitive abilities and the person may no longer be able to enact their contemporaneous values without support from others. We make a similar claim for the person's capacity to make sense of their circumstances, given the cognitive limits to their self-understanding. Sense-making requires additional support and scaffolding in dementia, and it will also be confined, simplified, and local in its scope. But the person still retains the need to make sense that is constitutive of agency and can be assisted to do so.

These arguments may not yet convince some philosophers that Mr. Q has sufficient self-understanding to justify our normative stance of respect for sense-making. They will say that Mr. Q's merely ostensible self-understanding implies that the agency he displays and the sense-making he engages in are a simulacrum of the real thing, deserving of kind support, but not respect. It is to this objection we now turn.

VI.2. Second objection: Marginal Agents, Respect, and the Principle of Beneficence. For the remnant identity cases it might be claimed that it is beneficence that does and should govern any decisions about whether or not to 'go along with' or facilitate activities arising from the outdated self-image. For imagine that the person's beliefs about their

⁴⁴ Sabat's case, mentioned above, provides a particularly enlightening contrast. Dr. B retained insight into his condition, and could reflect accurately on his experience despite the fact that, as he put it, "things get jumbled, and Alzheimer's gives me fragments" (Sabat, *Alzheimer's Disease and Dementia*, *op. cit.*, p 63). The disease, he said, was "constantly on my mind" (*Ibid.*, p. 58). Dr. B seems clearly to have a more robust self-understanding in virtue of this awareness of his predicament than what we have described as remnant identity cases. Illness insight is an epistemic signal that helps to intelligently interpret and filter experience.

⁴⁵ Contrast these cases with that of a person who develops a florid identity delusion such as the case described above of the person who believes they are a world-famous physicist. The deluded self is neither grounded in the person's own history, nor responsive to their circumstances. In such cases we respect the person's agency and their continuing and pressing need to make sense of the world, by *restoring* them to a true sense of who they are. For discussion of these and other cases where no such restoration is possible see Matthews and Kennett, "Respecting Agency in Dementia Care," *op. cit.*, pp. 124–27. (Thanks to an anonymous referee for pressing us on such cases.)

⁴⁶ We suggest, then, that the narratives Mr. Q and Martha construct to make sense of their circumstances meet Bortolotti's 'epistemic benefit' and 'no alternatives' conditions for epistemic innocence discussed above. See "The Epistemic Innocence of Motivated Delusions," *op. cit.*, p. 496.

⁴⁷ Agnieszka Jaworska, "Respecting the Margins of Agency: Alzheimer's Patients and the Capacity to Value," *Philosophy and Public Affairs*, XXVIII, 2 (1999): 105–38.

⁴⁸ *Ibid.*, p. 117.

surroundings or their role were distressing to them. Imagine that they came to believe they were still in a concentration camp, that their carers were guards, and that their fellow ‘inmates’ were being mistreated.⁴⁹ Their sense-making efforts here surely do not generate any requirement of respect, and that shows that beneficence is the appropriate governing principle in our interactions with marginal agents.

We do not deny that beneficence is an important moral consideration in such cases, but we argue that the *form* our beneficence takes should be governed by respect even in such cases. A person’s attempts at sense-making reveals and partly constitutes them as *agents*. This is what all agents share and what is owed respect. As such it applies to cases of marginal agency as well as to cases of robust agency. Indeed, marginal agency may properly demand more of us, since the person with diminished or impaired cognitive resources often requires additional support in their struggles to make sense of a confusing world. Recognising that marginal agents (including persons with dementia or other cognitive disabilities, persons suffering psychosis, and so on) share the same project of sense-making that we all have, and seeing them as striving for intelligibility, should block the tendency to see them as less, as unworthy or incapable of interpretation, and so to retreat to a mechanistic or objective stance towards them – as something to be “managed, handled, cured or trained” to use Strawson’s unfortunate phrasing (2008: 10).⁵⁰ Respecting agency is not, in our view, equivalent to respecting autonomy and the duty to do so does not therefore evaporate even in the case of agents who lack the capacity for autonomy.⁵¹

It is understandable, however, that this form of respect might be seen as proxy for, or recast as, beneficence. Respect for, and support for, sense-making, often requires a kind and degree of engagement with the other that mere truthfulness does not. It is a more demanding and intimate mode of respect for persons. It requires us to have a more thorough knowledge of the person, their history, and their perspective. It requires us to take them seriously. As a duty that emerges most clearly in professional and personal relationships it comes intermingled with concern for the other. Compare the doctor who delivers a life-changing diagnosis truthfully and objectively, with one who in addition takes the time to assist the patient to make sense of it, to understand what it means for her, and how it might fit into her life. The second is more beneficent to be sure but it is so largely in virtue of the support provided for the patient’s need to make this unwanted event intelligible. It is more beneficent *because* it is more respectful. Our claim is thus that recognizing and approaching all agents, including so-called marginal agents, as striving to

⁴⁹ For examples of this distressing phenomenon see Jan Wong (2002), [The return of the Auschwitz nightmare - The Globe and Mail](#). We thank a referee for directing us to this article.

⁵⁰ See Peter Strawson, “Freedom and Resentment,” in G. Watson, ed., *Free Will* (Oxford: Oxford University Press, 1982), pp. 1–25, at p. 10. For a useful discussion of the case of psychosis, see Sofia Jeppsson, “Psychosis and Intelligibility,” *Philosophy, Psychiatry & Psychology*, XXVIII, 3 (2021): 233–49. For a discussion challenging the Strawsonian conception of marginal agency including persons with mental illness see Jeanette Kennett, “Mental Disorder, Moral Agency, and the Self,” in Bonnie Steinbock, ed., *The Oxford Handbook of Bioethics*, (Oxford: Oxford University Press, 2007), pp. 90–113.

⁵¹ Traditional conceptions of autonomy are targeted by Jaworska in her discussion of valuing in Alzheimer’s Dementia. (See Jaworska, “Respecting the Margins of Agency,” *op. cit.*.) We believe our claims about the importance of sense-making are compatible with her account and provide an additional ground for taking seriously the values expressed by persons with AD. As signalled in fn.1 though, our focus is firmly on agency, rather than autonomy. In general, we prefer not to talk about degrees of autonomy in part because it might give the false impression that the duties of respect that we describe come in degrees that match the level of autonomy of the agent. (If it is autonomy that demands respect, then less autonomy equates to less respect.)

make sense of themselves and their environment, is both appropriately respectful and provides a superior framework for conceptualizing their best interests.

With this in hand let us reconsider the case of the distressed dementia resident. We first need to appreciate and understand the cause of their distress as located in their agential need to make sense of their surroundings, and something in their surroundings as having triggered a role or narrative fragment that allows them to do so. So, we might explore whether the distressing narrative is triggered by particular environments within the care home, and if so whether we can alter those environments, so as to fit the environment to the person's capacity to make sense of it, or avoid them.⁵² We might offer alternative appropriate interpretations: "You escaped the camp, you are in a safe place now" or offer material which stimulates happier memories, more positive roles, and happier narrative threads and routines – photos, recordings and the like. Understanding the person's history and values, and respecting the person's sense-making needs guides a response which is both respectful and beneficent.

Compare this with an approach which would seek to remove distress without considering its agential and biographical origins. Perhaps we could give the person a sedative which removes their distress by dampening their sense-making drive – by effectively subduing their agency. They become comfortably numb and passive.⁵³ While there may be situations – particularly in late-stage dementia – in which this is the best or only option for relieving distress, we argue that where agency and a self persists, albeit in a diminished and fragile form, this should not be our first option. In general, working out *how* to be appropriately beneficent requires taking account of the agential needs of a person. That is the reframing we urge. It does not mean that we do not take appropriate steps to soothe a confused person's distress. Often that distress will be alleviated if we can alleviate their confusion – that is, assist them to make sense of their situation. But as with anyone else sometimes we should just offer comfort and reassurance, as when we give a grieving friend a hug and tell them that we will be there for them. We do not deny the value of beneficence. We do not deny that there are times when immediate concern for welfare takes priority in our dealings with both robust and marginal agents. What we do deny is that beneficence can replace the mode of respect we have outlined in this paper, or that it provides the overarching moral framework to guide our interactions with marginal agents.

VII. CONCLUSION

The moral significance and status of persons demands respect. In this paper we have focused on the forms of respect owed to persons *as agents* in our personal interactions with them. We described the centrality and importance of sense-making for agency and argued that support for each other's sense-making efforts is a form of respect that is distinct from the requirement for truthfulness. Truthfulness is indeed an important mark of respect for others. Lying undermines agency and autonomy because a person placed in possession of the withheld facts may have chosen to act differently; moreover, it prevents the deceived

⁵² We describe some actual cases of this in Matthews and Kennett, "Respecting Agency in Dementia Care," *op. cit.*, p. 121.

⁵³ A referee has provided a further interesting possibility. What if, instead, we could give a pill that would remove the distressing belief without subduing the person's agency? If we would give the pill doesn't that show that we are motivated by beneficence, rather than respect? We do not think that providing medication that restores or preserves agency is at odds with a requirement of respect, so we do not think this case constitutes a counter example to our claim.

person's narrative from being truth-tracking and threatens the good of intimacy and the possibility of personal development. Normally truthfulness and support for sense-making go together but, as the case of gaslighting shows, attacks on a person's self-image and credibility can undermine their ability to make sense of the world, and themselves, and to exercise their agency in the light of their self-understanding, even where there is no active deception.

Recognition of, and respect for, another requires us to support and enable their agency and this includes support for sense-making. In cases where a zealous insistence on truthfulness could threaten rather than support agency and self-understanding, we need to consider what route is more respectful of the other in all of the circumstances. While we do not doubt that truthfulness should often win out in situations of conflict, the cases we have considered here suggest that a rigid insistence on the truth does not merely lead to social discomfort or awkwardness; it can be disrespectful of the other and sometimes profoundly so, in disregarding their perspective on the world and their efforts to make sense of it and themselves.

Macquarie University

Australian Catholic University

JEANETTE KENNETT

STEVE MATTHEWS