

Do You Mind Violating My Will? Revisiting and Asserting Autonomy

Eli Benjamin Israel

April 14, 2024

Paper Draft

Forthcoming in Routledge's "The Philosophy of Sexual Violence," edited by Georgi Gardiner and Micol Bez. Please cite published version.

Content warning: this paper discusses sexual activity, rape, and sexual trauma.

Abstract: In this paper, I discuss a subset of preferences in which a person wants the violation of desire they chose to make effective, such as in rape fantasies. I argue that such cases provide us with a unique insight into personal autonomy from a proceduralist standpoint. In its first part, I analyze some examples in light of Frankfurt's endorsement theory and argue that even when we cannot endorse a practical decision that involves being violated, we nonetheless regard those cases as instances of autonomy. Therefore, autonomy does not necessarily require endorsement. Instead, I propose that the nature of the relevant highest-order volition dictates the procedure that should be established in one's desire structure for its fulfillment. In the second part, I discuss how an agent may effectively consent to the violation of their decision by another person. Ordinary consent typically refers to actions but fails to communicate one's higher-order desires or commitments. To address this limitation, I propose a practical tool that accomplishes this by signaling shifts in the normative context where the agents are interacting.

*Jamie's Case.*¹ Jamie and Kit have been seeing each other for a while. Their relationship is primarily sexual. Nonetheless, they have a strong bond and care much about each other. They are generally concerned with each other's fantasies and desires and are willing to realize any consensual sexual scenario for one another. One night, in a moment of candor after making love, Jamie turns to Kit, grabs their face, and says, looking right into their eyes, "Kit, I want you to do something for me. One day, I want you to catch me off guard and rape me."²

If Jamie is accurately conveying their erotic fantasy to Kit, and if this fantasy truly represents a desire they want to realize, then this is not a request for having sex while pretending not to be interested, nor does it merely express a desire for an unusually aggressive sexual experience. In the most literal sense, the object of Jamie's desire is not even a specific action but the violation of their decision to refrain from sex. Kit, on the other hand, will only consider fulfilling Jamie's fantasy if it is consensual, which implies (and I am being deliberately minimalist here) that they act in accordance with what Jamie wants.

Of course, this is a fringe case, but one worthy of philosophical investigation for two reasons. First, from the standpoint of sexual ethics and action, Jamie's case may help us better understand how far we can stretch one's sexual agency (particularly in sexual relationships where control and power are ceded). The norms and limits of "our ability to explore and pursue our desires and control our sexual narrative" (Kukla, 2018, p. 72) or of our free

For discussions and comments on earlier versions, I thank Eugene Chislenko, David Enoch, Brian Hutler, Oded Na'aman, as well as the audiences at the Israeli Philosophical Association 2022 meeting, the Great Lakes Philosophy Conference, and the Temple University Work-in-Progress Series.

¹ I recognize that this is a hard-to-swallow example, but I believe it is necessary one for the purposes of this paper, so bear with me here. I also acknowledge that one may have worries about the association of such a fantasy with some problematic gender stereotypes. In order to avoid these distractions and allow us to focus on the philosophy, I am purposefully assigning gender-neutral names and pronouns to the characters (by no means indicating their non-binarity).

² "Rape" is defined in different ways in different contexts. In our case, Jamie does not desire full-fledged rape, but alludes to a quite common sexual *fantasy* that makes use of the term (but assumes some form of willingness embodied in her request).

capacity to be "knowledgeable and forthright about [our] sexual needs, desires and interests" (Cahill, 2016, p. 755) are topics of growing interest in recent literature,³ and it may be useful to investigate whether Jamie's desire is a coherent one to have and pursue (especially to the extent that it requires the consensual participation of another person).

Second, and perhaps most importantly, Jamie's case provides us with a unique insight into personal autonomy—This is the primary focus of this paper. The literature on personal autonomy largely distinguishes between procedural and substantive accounts. The latter proponents argue that autonomy implies some normative competence (Benson, 1991; Wolf, 1990) or is subject to moral commitments associated with our moral nature (Hill, 2012; Korsgaard, 1992). On the other hand, proceduralists, such as Harry Frankfurt (1971) and Gerald Dworkin (1970, 1988),⁴ defend a liberal and content-neutral approach, in which autonomy entails the agent's endorsement of the desires that move her actions. According to this model, the harmony in one's desire structure determines her autonomy as it reflects that the agent has the will she wants to have (Eyal, 2019).

Now, leaving aside the obvious moral discomfort of substantivists and its impact on their evaluation of Jamie's autonomy, I suggest that this case may present us with a subset of autonomous preferences that challenges the proceduralist endorsement view: While performing the relevant action is not necessarily contrary to their higher-order desires or commitments (and if you do not share this intuition, I hope to persuade you about this), if we focus on Jamie's desire structure, we will see that endorsement is not possible because they desire the very *frustration* of their decision to refrain from sex. As a liberal who strives to

³ See, for example, Kukla (2018) and Miller (2022).

⁴ I focus on Frankfurt and Dworkin because they represent the most influential account, but developments of their views may be found in Watson (1975), Christman (1991, 1993), and from a feminist perspective, in Friedman (Friedman, 2002) and Meyers (2014). Nonetheless, I focus on the Frankfurt-Dworkin model due to its prominence, and I believe that my use of it can be generalized to other proceduralist approaches as they primarily differ in respect to the mental state that is required to be in harmony with one's actions (desires, evaluative judgments, long-term plans, etc.) (Buss & Westlund, 2018).

"accommodate a diversity of desires and ways-of-life as autonomous" (Christman & Anderson, 2005, p. 3), I am curious whether a better understanding of this model could accommodate such a desire and what conditions must be met for its consensual performance. These are the questions I address in this paper.

I start, in §1, by framing our case in terms of Frankfurt's account and explaining how desires, will and consent will fit into my analysis.⁵ In §2, I discuss the peculiar character of second-order volitions in Jamie's and other similar examples, leading to a revised view that is still content-neutral but assumes that the relationships between different-order desires (endorsement or frustration) should be determined in light of the character of the relevant higher-order desires. With an appropriate account of autonomy at hand, in §3, I proceed to discuss whether Jamie can effectively consent to the violation of their decision by another person. Because ordinary consent refers to actions but fails to communicate one's higher-order desires or commitments, I propose a practical tool that accomplishes this by signaling shifts in the normative context set by the interacting agents.

§1 Setting the Grounds

Our case study is characterized by tensions within the agent and between the agent and their partner. The internal tension may be what causes the reader to feel uneasy when Jamie uses the term "rape" to describe her desired scenario—not only because of the emotional baggage this term carries but also because we naturally assume that rape is in inherent contradiction with desire. This tension is the exact object of our investigation from the proceduralist standpoint; however, to address it skillfully, we must fill in the details of what inner states

⁵ Frankfurt and Dworkin have developed independently quite similar theories, and thus, for the sake of simplicity, I chose to develop my argument through Frankfurt's lens.

conflict. I will do so by focusing on Frankfurt's model, framing it as a particular type of conflict between different-order desires.

Frankfurt's conception of autonomy lies in the distinction between first-order desires—which are desires to X, where X=action—and second-order desires—which are desires about other desires and represent what one wants to want. Furthermore, he distinguishes between mere second-order desires and second-order volitions. The latter is a subtype of the former that aims at shaping the person's will. These are desires that, in order to be realized, must be reflected in one's actions, and they ultimately reflect the agent's deepest commitments. For this reason, Frankfurt argues, an autonomous action must be consistent with the agent's higher-order volitions (1971, p. 10).

Conflicts between different-order desires are common phenomena—people have multiple first-order desires, and if only some of them align with their second-order volitions, then the rest is in conflict. Nonetheless, when these conflicts arise between the agent's second-order volitions and a first-order desire that constitutes her *will*, they threaten her autonomy.

Frankfurt's notion of 'will' is that of "an *effective desire*—one that moves (or will or *would move*) a person all the way to action." (Frankfurt, 1971, p. 8).⁶ Thus, according to him, the will is akin to one's choice or practical decision: it is identical to the first-order desire that takes the upper hand and ultimately determines her intended action.

Most conflicts between second-order volitions and one's will result in non-autonomous actions. Frankfurt's Unwilling Addict is a paradigmatic example of this. It tells the story of a drug addict who has two competing first-order desires: to use the drug and abstain from it. Furthermore, he has a second-order volition directed toward the latter—he wants the desire to

⁶ I emphasize "would move" to convey that Frankfurt does not require the effective desire to actually produce an action, but only for the desire to effectively determine one's intention to act.

abstain from using the drug to become his will—but, unfortunately, his desire to use the drug is too strong and ends up constituting his will and determining his action. The addict is driven by his physiological condition rather than by his decision. He cannot endorse the desire that determined his action; therefore, he is not autonomous. Jamie's case, on the other hand, is quite different as the object of their second-order volition is not a first-order desire that fails to constitute their will but the very frustration of their will (in a sexual context, by a specific person).⁷

To break this down, Jamie requests a scenario in which: (i) they have a first-order desire to refrain from sex; (ii) the latter takes precedence over their competing desires, becomes their will, and they are motivated all the way to action to refrain from sex; and (iii) Kit violates their will by having sex with them. In this way, Jamie's will becomes effective on their end, meaning it *would* move them to action, but Kit, due to Jamie's request, prevents the action from actually occurring.

As we progress to the third stage, which represents the tension between the agent and their partner, we must discuss the consent Jamie grants to Kit. Now, the conception of consent associated with proceduralist views is concerned with negative freedom, avoiding any type of external coercion or manipulation and aiming for authenticity, as opposed to substantivist views, which aim for self-control (cf. Miller, 2022, p. 248; Ursin, 2009, p. 19).

This characterization, however, is insufficient because the case we are discussing is not action-oriented. Jamie's consent is not for Kit to perform a specific action but rather to have an impact on their will. Therefore, we must broaden the scope of consent so that it becomes a means by which an individual not only addresses actions done to her but also "reorders her

⁷ This is also consistent with the working definition used in empirical studies on the nature and prevalence of rape fantasies that frame those as "waking fantasies that depict the use of force or incapacitation by a male or female *against the will of the self-character* in the fantasy" (Bivona & Critelli, 2008, pp. 3–4, emphasis mine), instead of referring to those as consent violations.

normative context" (Higgins, 2004, p. 98). This view is somewhat similar to Joseph Raz's account, which defines consent as "acts believed to change the normative situation because they are performed with that belief" (Raz, 1988, p. 84).

The reason for adopting this broadened view of consent lies in the very nature of cases like Jamie's. It addresses one's set of beliefs and desires rather than specific acts and their permissibility. In this way, it also aligns with the content-neutral character of proceduralist views. The agent does not commit to any specific acts but rather agrees with their partner on a normative context that allows a variety of acts that would not be permissible in ordinary settings (Higgins, 2004, p. 100). This view reflects well that Jamie's request is not directed at any specific action, but instead, it allows Kit to engage with them considering a scenario in which violating their will is a legitimate action.⁸ Furthermore, it establishes a proper framework for the partners' communication in situations where no speech act aimed at preventing Kit's action can be effective.⁹

With this, I believe we have a decent conceptual framework for approaching the case.

§2 Revisiting the Character of Second-Order Volitions

As noted in the previous section, what is unique in the analysis of Jamie's case is that we face a second-order volition that requires the violation of the agent's will to be fulfilled. This constitutes a challenge to accounts like Frankfurt's that assume the endorsement as the

⁸ As Wertheimer argues, claiming that an act or behavior is consensual does not imply that it is also morally permissible. Consent is morally transformative in the sense that doing things to other people is a wrong that we avoid when we give consent. As a result, consent gives us some reason (though not conclusive) to believe that an act is morally permissible. (Wertheimer, 1996, p. 90). So, if it was not clear before, I would like to clarify that I am not discussing the overall permissibility of carrying out the act Jamie requests, but only its consensual status as it relates to their agency.

⁹ The question of how consent looks like when aimed at the normative context is a puzzling one which I address in the third section of this essay.

relationship to be established throughout one's desire structure for an action to be autonomous.¹⁰

Endorsement is indeed important, and we might even say that, in most cases, it correctly reflects the desire structure that results in autonomous actions. When I say, "I wish I were more motivated to exercise," or "I want to be the kind of person that wants to exercise," I am expressing a second-order volition that must align with my will to be fulfilled. Looking up my desire structure, it also resonates with my desire to be in better health, which, in turn, resonates with my desire to have a long and pleasant life and so on throughout my deepest (and authentic, I may say) commitments.

However, I want to argue that the requirement for alignment or endorsement should not be assumed to be the only way a second-order volition shapes the agent's will. Frankfurt's identification of the will with a first-order desire that becomes effective and his definition of second-order volitions as desires aimed at shaping one's will imply only that these volitions are satisfied when they lead to *corresponding* actions. Namely, actions that reflect the content of the relevant higher-order desire, with no specific requirement for endorsement throughout the agent's desire structure. This is not to say that we must reject procedural autonomy, but rather that the procedure or formal (content-neutral) requirement should be determined by what the higher-order desire needs to be fulfilled. According to this suggestion, the relevant volition does dictate not only the action's content but also the inner procedure to obtain it.

I think it would be useful at this point to look at an analogous case from the realm of moral philosophy:

¹⁰ The harmony or endorsement is not necessary only between one's will and her second-order volitions alone, but it is important also for the latter to stand in harmony with desires that stand higher in the structure. This constitutes a common point of criticism of Frankfurt, for the potential regress of higher-order desires (for more on this criticism and on Frankfurt's responses to it, see, Cummings & Roskies, 2020, pp. 424–425) .

Peter's Case. Peter is a moral expert. By *moral expert*, I mean he is very good at making correct moral judgments for the right reasons.¹¹ He was, however, unfortunate enough to be born as a highly akratic person —whenever his passions tempt him to do the wrong thing, he does it while knowing he ought not. His weak will stems from how unbearable the frustration of not fulfilling his passions is to him. Now suppose Peter decides to work on himself, and he does so by addressing this unbearable feeling of frustration. He wants to put himself in tempting situations where he will be moved to act immorally and asks his devoted friend, Mary, to intervene and prevent the actions from taking place. So despite his moral commitment to protecting the lives of animals, he begins to attend meat festivals across the country on a regular basis. He sees all those briskets, succulent lamb chops, rare steaks, and juicy dripping sausages (which he is very fond of), and he brings himself to these tempting situations just to experience and overcome the frustration that comes from not sinking his teeth into a sizzling ribeye steak.¹²

When Peter performs this exposure therapy, his will is not directed towards any specific action. Instead, he is working at changing his practical decision process, and he primarily desires the frustration of his passion-driven will. He desires so not because he wants to refrain from eating animal parts but because he intends to alter patterns that diverge his practical decisions from the right path.

¹¹ There is no agreement in the literature about the conditions for moral expertise. While some philosophers regard that a moral expert must be virtuous, namely one whose actions are morally good (see, Burch, 1974 and Szabados, 1978), others claim a moral expert must only be successful in producing correct moral judgments for the right reasons (see, McConnell, 1984) . For the sake of simplicity in this example, I will settle for the latter sense.

¹² This example is inspired by Frankfurt's (1971) Unwilling Addict, with the primary difference being that the addict is not in a reflective state aimed at changing his practical decision process, but rather is simply aimed at favoring the desire to refrain from taking the drug (which requires endorsement to be fulfilled).

This clearly constitutes a second-order volition, as second-order volitions are not limited to the approval of lower-order desires. Their object must be a first-order desire, but they may stand in various relationships with those.¹³ Therefore, they can also relate to the agent's will, aspiring to violate it. Furthermore, this case appears paradigmatically autonomous, as this desire resonates with his moral commitment, which we assume is pure because he accesses correct moral judgments for the right reasons.

By drawing a parallel between Peter's and Jamie's cases, we can see that both protagonists have a second-order desire to violate their will. In Peter's case, it is clear that he is acting autonomously in a way that will eventually lead to the reflection of his moral knowledge into his behavior. As a matter of fact, he is acting contrary to what we saw as an example of non-autonomous action in Frankfurt's unwilling addict example. That being so, it is wrong to assume that the exposure therapy he designed for himself is not autonomous simply because of the formal setting of his desire structure. The violation of his will is the only way for his volition to shape his will accordingly.¹⁴ Similarly, it would be wrong to rule out Jamie's desire as possibly autonomous simply because endorsement of their will is contrary to their volition.

The fact that it is only possible to meet Peter and Jamie's deepest commitments if we accept a discordance between the person's will and a relevant higher-order volition brings us to reject endorsement throughout one's structure as a necessary demand for autonomous actions. It is more plausible to say that we should be concerned with whether the relevant higher-order

¹³ Here I reject a common misconception about Frankfurt's theory. Second-order desires are commonly interpreted as if they could only approve or reject other desires. In other words, while first order desires may have a variety of forms and relate to their object in many ways, second-order desires are often understood merely as "wanting to want X" or "wanting not to want X". This, however, is a very odd and unjustified limitation. If different-order desires are only distinguished by their object, then they must accommodate not only the form of "A wants X" but also that "A wants to do Y with/to/for X."

¹⁴One may suggest that what makes Peter's desire autonomous is not his desire to train his will (i.e., to have his desire for meat frustrated) but his previous choice to have this desire. The latter may be understood as a desire to desire to have his urge for meat frustrated. This might well be true, but it is unimportant for our purposes as it would still preserve the discordance between desires in that same structure..

volition is, in fact, autonomously determined (authentic); and what kind of relationship it imposes on the desires under it in one's desire structure. We may accept that most volitions require endorsement and conformity, but this is not true in all cases.

Thus, I suggest turning the discussion from the structure of desires to the character of the relevant higher-order volition that moves a particular action. Peter is an autonomous agent because he decided to frustrate his will out of authentic moral concerns. The performance of Jamie's fantasy, on the other hand, would be an autonomous act only if their relevant higher-order volition is truly authentic, non-alienated, and not the result of external conditions. In that case, the only way to determine whether the act is autonomous is to delve deeply into their motivations. The degree to which Jamie would be more or less autonomous would depend on various factors, including the intrusive nature of their motive, their awareness of it, and even their attitude toward it during the action.¹⁵

In order to account for the authenticity of this desire, we may speculate about Jamie's psychology. On the one hand, we can attribute their desire to some innate masochism or a spontaneous desire to lose control. Assuming they come from Jamie and are not influenced by external factors, it seems safe to say they are authentic, and acting in accordance with those desires would constitute an autonomous action. On the other hand, their fantasy could reflect a desire to safely access a previous experience of assault "under lab conditions." In such a case, acting on this desire would not be taken as a sufficiently autonomous action because the desire has its origins in an external factor (the assault).¹⁶ Each of these psychological explanations has an impact on the authenticity of Jamie's rape fantasy, reflecting directly on their autonomy (if the corresponding actions are taken). Nevertheless,

¹⁵ The latter refers to instances of reclaiming acts, such as when a minority group adopts derogatory language/acts/attitudes directed at them and reclaims them as part of their identity.

¹⁶This account is much present in the literature about adaptive preferences, as part of the discussion on "autonomy deficit" (Stoljar, 2014). This example and analysis are based on the self-starving woman case described in David Enoch's recent paper (2020, pp. 183–185).

leaving those speculations aside, there is no reason to assume that acting on such a desire would not correspond with Jamie's relevant higher-order desire, nor that the latter is in intrinsic dissonance with their personhood based solely on the formal conditions of Jamie's desire structure.¹⁷

With this, we conclude the first part of the essay, which addressed the agent's internal conflict as reflected in her autonomous status. I argued that cases like Jamie's and Peter's, in which a person freely desires a violation of her will, demonstrate that we must revise Frankfurt's view and take the formal relationship between desires to be contingent on the character of the relevant higher-order volition. Although endorsement is the most common relation motives impose on the agent's desire structure, it should not be interpreted as a general rule. The relevant procedural/formal feature of autonomy in certain exceptional cases, such as ours, is the disparity between one's second-order volitions and will.

§3 Consent and Assertions of Autonomy

So far, I defended that one's second-order volition to violate their own will is possibly autonomous. Assuming instances where this is the case, we must inquire how consent may be practiced to voice such a volition. As I briefly state in §1, consent is usually considered an action-oriented speech act that allows others to act on us or to do things to us. Therefore, it can, at best, reflect our will. I say "at best" because we may consent to things we do not desire, of course, but when directed positively to desires, action-oriented consent can only reflect action-oriented desires (i.e., first-order desires).

¹⁷ In fact, if to say anything empirical about this, research actually shows that individuals that have rape fantasies generally tend to have less sexual guilt and more openness to their true sexual desires, making these fantasies a "generally open, positive, unrestrictive, and relatively guilt-free expression of one's sexuality" (see, Strassberg & Locker, 1998, p. 413; cf. Bivona & Critelli, 2008, p. 64). This is not to say that these fantasies are necessarily autonomous. These empirical findings only support that these desires may freely pertain to the agent.

This limitation may hinder one's autonomy in situations where their will and its expressions are set to be contradicted by another person. In Jamie's case, their request from Kit communicates their second-order volition and establishes a normative context in which Kit may violate Jamie's will. Setting this context is a form of consent, according to the broad definition of consent I presented in §1. Still, it is insufficient to ensure the experience's consensual character because the conditions that established this consent (Jamie's second-order volition) may change. The fact that Jamie has this volition now does not imply that they will have it in the future, and there must be a way to communicate such a shift.

The same holds for Peter. There is an important sense in which Peter must be able to communicate to Mary that there were changes in the second-order volition that initially led him to request her to prevent him from eating meat.¹⁸ David Enoch argued for the necessity of this possibility in order to preserve one's autonomy with the following example:

Suppose that my daughter, out of concern for my health, will sometimes hold on to the salt at the dinner table and refuse to pass it along, even when I ask her. I appreciate the gesture and the sincere concern, of course. But sometimes, I insist. I can explain that I understand the health issues, but that it's my body, and my life, and that I want the salt, please. I can, it is natural to say, *assert my autonomy*. At least in some such cases, if I so insist and my daughter still refuses to let me have the salt, she is offending against my autonomy. (Enoch, 2017, p. 31, emphasis in original)

In Enoch's example, he asserts his autonomy by reassuring his daughter that he is aware of the health risks and still wants the salt despite them. This is possible by verbally reaffirming his will because, even though his daughter is aware of his greater health commitments, if she

¹⁸ This does not contradict his moral obligation to refrain from eating meat. The important question for us is not what is morally correct, but what Peter is free to do as an autonomous being. This is another instance where the neutrality of proceduralist views is apparent. See, Enoch (2002) for more on the right to violate one's duty.

regards him as an autonomous being, she still should regard his affirmation as an expression of his autonomy. It is not practically necessary for him to maintain his health commitments at all times, and he is free to suspend them for any reason. Having the final say over his actions is a necessary feature of his autonomy, which ultimately makes him morally accountable for his actions.

We may support this point by alluding to Derek Parfit's famous story of the Russian Nobleman. It tells the story of a young Russian socialist who plans to donate his lands to peasants when the time comes. Due to his fear that his idealism will fade, he signs a legal document that automatically gives away his lands and asks his wife to promise not to revoke it even if he asks her to.¹⁹ As Christine Korsgaard argues in her analysis of the story, the Russian socialist is wrong in dismissing the autonomy of his future-self, and in doing so, he also wrongs his wife. She argues that if we do not hold others, and ourselves at other times, responsible for their actions, we do not regard them as moral and rational agents, and thus, neither as autonomous beings (Korsgaard, 1996, pp. 206–207). It is necessary for the preservation of the person's autonomy that they are able to communicate what they want at all times, and shape their actions accordingly.

However, in the cases we are interested it is more difficult to assert the agent's autonomy as the conditions for consent change. Jamie's verbal reaffirmation of what they want (as Enoch did in the salt case) will not suffice because it is exactly the violation of their will that the agent desired from the start. If they say, "I do not want you to do X to me," and Kit is committed to the normative context they have established together, then X is precisely what Kit should do.

¹⁹ See, Parfit (1973, p. 145ff, 1987, pp. 327–328).

For this reason, to provide a comprehensive account of the consensual realization of this subtype of volitions, we must find a method for the agent to communicate shifts in the conditions for consent that circumvent their will. This would allow Jamie and Kit to confidently carry out her fantasy, as well as Peter to assert his autonomy if he changes his mind about his practical decision-making process or the moral permissibility of eating animals.

I want to propose a solution that comes from an unconventional direction, but that is not all the way surprising, given the leading case of this essay. It is inspired by a particular dynamics of safewords, and I believe it will not only work well in the cases I am trying to make sense of, but it also has the potential to open the floodgate to interesting discussions and applications of this tool. In fact, there is no substantial philosophical treatment of safewords other than that of Quill Kukla (writing as Rebecca Kukla) in "That's What She Said: The Language of Sexual Negotiation" (2018). For this reason, my account will be extensively based on theirs (Kukla's).

Kukla argues that "the use of safewords is a fascinating type of speech act with multilayered illocutionary force and a complex set of governing conventions" (Kukla, 2018, p. 87). While they are primarily used in kink-sexual contexts, they claim that there is no reason for limiting it to this domain and defends its use as a standard practice in all forms of sexual relations. I want to go even further, incorporating it into the broader discussion of autonomy and consent.

Following Williams et al.'s (2014) distinctions in levels of consent, Kukla argues that "safewords are part of what we establish during *scene consent*, and they are tools that aid with *deep consent*" (Kukla, 2018, p. 88, my emphasis). *Scene consent* refers to communicating what we want and allows to happen to us in the middle of an act or "in-scene;" *Deep consent* refers to a general awareness of how comfortable a person feels with

the situation being carried out.²⁰ In the same way, the method I want to propose signals the status of our second-order volitions to the partner in *scene consent*, making her aware of the normative context we are currently in and contributing to her general awareness not only of what she is allowed to do but also of their partner's inner state.

When we think of safewords, we usually think of a word that participants agree on as a way to end sexual activity.²¹ These are words that stand out in a sexual context and, therefore, cannot be confused with something people may say in the common erotic speech. To some, safewords must also be able to "present themselves" in a psychologically salient way for the individual to draw her attention to the partner's discomfort immediately. It may be a word with a specific and prominent color or smell, such as "cinnamon," as many of us learned from Homer Simpson in the second episode of *The Simpsons'* twenty-fourth season.

The type of safewords I want us to consider is unique because it is not action-oriented. It does not command the situation itself (at least not necessarily) but rather draws the other person's attention to the pre-established normative context participants have agreed upon. Let us illustrate how it works, following the "traffic light" safeword system.²²

In this system, saying "Red " naturally signals that the person wants the sexual act to end. This, however, is not much different from the act of asserting one's autonomy as we saw it in Enoch's salt case. Saying "Red" is simply a way of stating unequivocally that one wants things to come to a halt. Saying "Yellow," on the other hand, represents a more interesting and nuanced type of speech act. Kukla initially describes it as "a way of indicating discomfort

²⁰ Kukla points out "it is clear that the language of consent is being asked to do too much work here, but the distinctions they are drawing is important." (2018, p. 87). If I understand them correctly, the language of consent is being stretched here in the same way I do in the first section, amplifying the notion of consent from the merely action-oriented form to one that takes the normative context created into account.

²¹ Some find safewords to be only capable of voiding consent or being an opt-out procedure. See, for example, Rocha (2016).

²² See, Kukla (2018, p. 87). For accounts of this system in popular platforms, see, (Kukla, 2019; Pullar, 2022)

or wariness and calling on the other person to ease off and be on the lookout for signs that the speaker wants to alter or stop an activity" (2018, p. 87).

The exact meaning of each color and how the partner should respond is up to each couple to agree on and set the rules in advance. However, with that same flexibility, we may leave out the action-oriented elements of 'yellow,' such as "easing off," and focus on the indication of wariness and the request for the other person to be on the lookout for signs of discomfort.

These signs could be related to how the person feels about the situation, her sense of safety at the time, or a shift in her mindset that suddenly does not feel so good about the normative context they have established. Using the 'yellow' safeword does not imply a transgression or some exaggeration in force on the partners' part.

The discomfort stems from the actions that occur or could occur within the established normative context. It is about what happens or could happen to the person, not about the norms apart from their practical implications. As a result, the call for the partner to be on the lookout for signs of discomfort indicates that the actions implied in the current normative context do not feel quite right, and thus the norms governing that situation should be reconsidered. Because the normative context was established as a result of the person's second-order volition, we can say that their practical discomfort results from a change in that initial state.

This suggestion can be taken as a practical version of Kukla's speech-oriented account, which sees safewords as tools for shaping the norms that lead participants' discourse in sexual encounters. Following Nicolle Wyatt's (2009) conception of second-order conventions, Kukla argues that safewords establish such conventions between participants, shaping how they

interpret each others' words and intentions.²³ That is, by agreeing on a safeword, partners bracket the original meanings of particular words and establish new norms to which they should respond and react.

When specifically referring to 'yellow,' Kukla says it "functions as not so much an order as a direction of *attention*, along with a call to shift gears a bit." (2018, p. 89, emphasis mine). In this manner, the salience feature I mentioned while explaining the general use of safewords is invoked. This is especially important because 'yellow' functions by making the partner attend to all sorts of pre-defined conventions rather than commanding the action being performed. Having a safeword with prominent presentational content, such as the color "yellow," may involuntarily guide one's attention, forcing them to move from one level of speech to another.²⁴

According to Kukla, "the skilled use of safewords requires participants to move between three levels of speech." We may consider this argument as we go back to the practical domain. In a sexual scene, there is usually an erotic speech level, which is often playful, metaphorical and staged; an ordinary speech level; and a second-order or metaspeech level, which establishes the rules and boundaries of discourse. Safewords belong to the latter level and are tools for communicating how we want our words to be interpreted.

In my practical version of this argument, 'yellow' functions as a speech act that reorders the normative context in which actions occur rather than reordering the norms of speech. When

²³ Wyatt discusses second-order conventions in the context of silencing acts, of the kind that interest MacKinnon (1987) and Langton (1993) are interested at. For example, when a woman says "no," and her partner does not interpret it as a refusal because there is a more salient interpretation that portrays her as choosing to be there, and that interpretation is shaped by second-order conventions set in our society.

²⁴ Watzl, in his comprehensive analysis of attention, ascribes psychological salience a necessary motivational component, which he refers to as a felt motivational impact. This is based on the experience we have of having our attention drawn to a salient object, as a fire alarm, instead of actively deciding to attend to it. In the same way - following Husserl and some contemporary phenomenologists, such as Chudnoff (2012) - I suggest the quite plausible claim that words we use also have a presentational content that may be salient and passively guide our attention. For more on Watzl's account of psychological salience, see , chap. 6 of "Structuring Mind: The Nature of Attention and How it Shapes Consciousness" (2017).

agents design a scenario (again, not necessarily a sexual one), they do not only discuss how their speech should be interpreted but also what they allow as legitimate actions within the scene. In this way, they create a particular normative context distinct from the one we usually live with. With this additional context, we have three practical levels: the temporary one they have designed, the ordinary one, and a meta-level in which second-order practical conventions are set. In the same way that 'speech-yellow' is expressed in an erotic context but draws the partner's attention to a possible return to ordinary speech, 'practical-yellow' may signal a parallel call that makes the partner aware that something in the normative context may have changed. It indicates a possible change in the normative statuses introduced by a previous illocution and that they might have to be reconsidered or completely voided if so the agent indicates. Safewords of this sort are speech acts that "undo" things, in Laura Caponetto's (2020) sense: they do not say anything positive about the normative context but merely signal the need for reconsideration.

If a participant in an erotic scenario says, "stop! I cannot go on with this," after saying "yellow" a minute before, the request for stopping should be treated as ordinary speech rather than erotic-level speech. Similarly, Jamie's 'practical-yellow' call (or a call of whatever word participants have agreed on with the same function) should indicate to Kit that what was permissive under their particular normative context should be suspended. When Jamie uses the 'practical-yellow' safeword, she indicates that the desire to have their will violated *may* not be present at the time and that Kit should be aware of it and watch for signs of unwillingness on her part.²⁵ In Peter's case, if he uses a 'practical-yellow' safeword when entering a barbecue fair, Mary should be aware that something in the normative context of his actions might have changed. If he decides to eat a steak, she should consider that as a

²⁵ I say it only may not be present because there is some vagueness embedded in this tool that cannot be dismissed. The use of this safeword only indicated a suspension of the norms previously agreed on, but they do not state something positive about the consent granted in the formation of the normative context.

possible expression of his (wrong) moral judgment rather than his akratic behavior.

Therefore, she should not prevent him from eating (as far as she is concerned about his request and not about preventing him from wrongdoing).

Introducing 'practical-yellow' safewords in this way, as a common discursive practice, complicates interpersonal relations. They require the prior establishment of rules and conventions exclusive to the participants and cannot be included in general discourse. Nonetheless, they enable us to engage in various actions that would otherwise be too dangerous to undertake. Without such a tool, it is rationally impossible to freely delegate any part of our practical reasoning, sometimes against our first-order judgments.

§4 Conclusion

This paper began by examining the unusual case of Jamie's rape fantasy, revealing a unique subset of preferences characterized by the volition to violate one's own will. This category revealed that we must revise Frankfurt's theory and take the endorsement of the will to be a contingent feature of second-order volitions.

When we focused on cases of self-violation, we discovered that their requirement for external compliance to be realized drags us to formulate a method for asserting one's autonomy or communicating changes in the normative context. This was an opportunity to introduce a novel practical tool: 'practical-yellow' safewords. This tool cannot pinpoint specific changes in the conditions that led the speaker to set a particular normative context. However, it does have the capacity to make the other participants wary and help them act with sensitivity to the partner's volitions.

I want to address three main objections to this argument. First, one might question the very existence of the subset of preferences/desires I am discussing and accuse me of incorrectly

framing the cases. They would deny that one could ever desire to be genuinely raped or that one could want to frustrate his desire for meat. At best, Jamie could have a false impression about what she wants (perhaps due to patriarchal conventions), and Peter would want not to have a desire for meat at all. To such an objection, I would first argue that it is presumptuous to assume that Jamie's desire cannot be authentic by definition. Second, it is unpragmatic to assume that Peter simply does not want to desire what is immoral. Peter understands that his tastes and desires are beyond his control, and he must want to overcome them to change his behavior (otherwise, he would be trapped in misery with no true prospect of improving his behavior).

The second objection addresses the fact that I conclude from Jamie's and Peter's cases that endorsement is not a required formal demand. Instead, it could be argued that the harmony required for a specific desire to be autonomous comes only from that desire and higher up in the hierarchy. This is an intriguing position to take if we accept an internalist approach to evaluating one's autonomy by relying on the autonomous status of one's desires and preferences. However, I find it difficult to justify dropping the endorsement requirement as a result of that desire. It appears that it would have an impact on the possibility of autonomously realizing what we consider to be an autonomous desire, because once non-endorsement occurs in one's desire structure, it would affect the desires and actions below that point.

The third objection is that we cannot make the 'yellow'-type safeword act only as a discursive tool and not an order directed towards the partner. This might be true, and a partner worthy of our trust might see those alerts about the normative context as an order to refrain from any action that would be wrong outside the particular set of rules previously agreed upon. Still, I do not think it counts against my suggestion: the fact that this discursive tool has immediate practical implications is good. The only reason we limit them to their discursive capacity is

that that is the only level at which the speaker can communicate about his normative commitments in no uncertain terms.

There is still much to be done in the pragmatics of 'yellow' safewords, but I believe I have made the case that it is a promising practical tool outside of its original home in the kink-sexual context, and that philosophers should pay more attention to and work on different uses for it. The possibility of extending the language of autonomy to the rules of speech makes a compelling case for this tool that should not be ignored.

References

- Benson, P. (1991). Autonomy and Oppressive Socialization. *Social Theory and Practice*, 17(3), 385–408. <https://doi.org/10.5840/soctheorpract199117319>
- Bivona, J. M., & Critelli, J. W. (2008). Women's Erotic Rape Fantasies: An Evaluation of Theory and Research. *The Journal of Sex Research*, 45(1), 57–70. <https://doi.org/10.1080/00224490701808191>
- Burch, R. W. (1974). Are There Moral Experts? *The Monist*, 58(4), 646–658. <https://doi.org/10.5840/monist197458444>
- Buss, S., & Westlund, A. (2018). Personal Autonomy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2018/entries/personal-autonomy/>
- Cahill, A. J. (2016). Unjust Sex Vs. Rape. *Hypatia*, 31(4), 746–761. <https://doi.org/10.1111/hypa.12294>
- Caponetto, L. (2020). Undoing things with words. *Synthese*, 197(6), 2399–2414. <https://doi.org/10.1007/s11229-018-1805-9>

- Christman, J. (1991). Autonomy and Personal History. *Canadian Journal of Philosophy*, 21(1), 1–24.
- Christman, J. (1993). Defending Historical Autonomy: A Reply to Professor Mele. *Canadian Journal of Philosophy*, 23(2), 281–289.
- Christman, J., & Anderson, J. (Eds.). (2005). *Autonomy and the Challenges to Liberalism: New Essays*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511610325>
- Chudnoff, E. (2012). Presentational Phenomenology. In S. Miguens & G. Preyer (Eds.), *Consciousness and Subjectivity* (pp. 51–72). Ontos Verlag.
- Cummings, R., & Roskies, A. L. (2020). Frankfurt and the Problem of Self-Control. In A. R. Mele (Ed.), *Surrounding Self-Control* (p. 0). Oxford University Press.
<https://doi.org/10.1093/oso/9780197500941.003.0022>
- Dworkin, G. (1970). Acting Freely. *Noûs*, 4(4), 367–383. <https://doi.org/10.2307/2214680>
- Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511625206>
- Enoch, D. (2002). A Right to Violate One’s Duty. *Law and Philosophy*, 21(4/5), 355–384.
<https://doi.org/10.2307/3505056>
- Enoch, D. (2017). Hypothetical Consent and the Value(s) of Autonomy. *Ethics*, 128(1), 6–36.
<https://doi.org/10.1086/692939>
- Enoch, D. (2020). False Consciousness for Liberals, Part I: Consent, Autonomy, and Adaptive Preferences. *The Philosophical Review*, 129(2), 159–210.
<https://doi.org/10.1215/00318108-8012836>
- Eyal, N. (2019). Informed Consent. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2019). Metaphysics Research Lab, Stanford University.
<https://plato.stanford.edu/archives/spr2019/entries/informed-consent/>

- Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 68(1), 5. <https://doi.org/10.2307/2024717>
- Friedman, M. (2002). *Autonomy, Gender, Politics*. Oxford University Press.
- Higgins, R. C. A. (2004). Consent, Residence, and the Democratic Voice. In R. C. A. Higgins (Ed.), *The Moral Limits of Law: Obedience, Respect, and Legitimacy* (p. 0). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199265671.003.0003>
- Hill, T. E. (2012). Kantian autonomy and contemporary ideas of autonomy. In O. Sensen (Ed.), *Kant on Moral Autonomy* (pp. 15–31). Cambridge University Press. <https://doi.org/10.1017/CBO9780511792489.003>
- Korsgaard, C. M. (1992). Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations. *Philosophical Perspectives*, 6, 305. <https://doi.org/10.2307/2214250>
- Korsgaard, C. M. (1996). *Creating the kingdom of ends*. Cambridge University Press.
- Kukla, R. (2018). That’s What She Said: The Language of Sexual Negotiation. *Ethics*, 129(1), 70–97. <https://doi.org/10.1086/698733>
- Kukla, R. (2019). Consent and refusal are not the only talking points in sex. *Aeon*. <https://aeon.co/essays/consent-and-refusal-are-not-the-only-talking-points-in-sex>
- Langton, R. (1993). Speech Acts and Unspeakable Acts. *Philosophy and Public Affairs*, 22(4), 293–330.
- MacKinnon, C. A. (1987). Francis Biddle’s sister. In *Feminism Unmodified* (pp. 163–197). Cambridge University Press.
- McConnell, T. C. (1984). Objectivity and Moral Expertise*. *Canadian Journal of Philosophy*, 14(2), 193–216. <https://doi.org/10.1080/00455091.1984.10716378>

- Meyers, D. T. (2014). The Feminist Debate Over Values in Autonomy Theory? In M. Piper & A. Veltman (Eds.), *Autonomy, Oppression, and Gender* (pp. 114–140). oxford university press.
- Miller, S. (2022). Sexual Autonomy and Sexual Consent. In D. Boonin (Ed.), *The Palgrave Handbook of Sexual Ethics* (pp. 247–270). Cham, Switzerland: Springer.
- Parfit, D. (1973). Later selves and moral principles. In *Philosophy and Personal Relations*. Routledge.
- Parfit, D. (1987). *Reasons and persons* (1. issued in paperback (with corr.), reprinted with further corr). Clarendon Press.
- Pullar, J. (2022). How To Ask For Sexual Consent Without Ruining The Mood. *Elle Australia*. <https://www.elle.com.au/culture/how-to-ask-for-consent-sex-26943>
- Raz, J. (1988). *The Morality of Freedom* (First Paperback Edition). Oxford University Press.
- Rocha, J. (2016). Aggressive Hook Ups: Modeling Aggressive Casual Sex on BDSM for Moral Permissibility. *Res Publica*, 22(2), 173–192. <https://doi.org/10.1007/s11158-015-9291-0>
- Stoljar, N. (2014). Autonomy and Adaptive Preference Formation. In A. Veltman & M. Piper (Eds.), *Autonomy, Oppression, and Gender* (pp. 227–252). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199969104.003.0011>
- Szabados, B. (1978). On “Moral Expertise.” *Canadian Journal of Philosophy*, 8(1), 117–129. <https://doi.org/10.1080/00455091.1978.10716212>
- Ursin, L. Ø. (2009). Personal autonomy and informed consent. *Medicine, Health Care and Philosophy*, 12(1), 17–24. <https://doi.org/10.1007/s11019-008-9144-0>
- Watson, G. (1975). Free Agency. *Journal of Philosophy*, 72(April), 205–220. <https://doi.org/10.2307/2024703>

- Watzl, S. (2017). *Structuring Mind: The Nature of Attention and how it Shapes Consciousness*. Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199658428.001.0001>
- Wertheimer, A. (1996). Consent and Sexual Relations. *Legal Theory*, 2(2), 89–112.
<https://doi.org/10.1017/s1352325200000410>
- Williams, D. J., Thomas, J. N., Prior, E. E., & Christensen, M. C. (2014). From “SSC” and “RACK” to the “4Cs”: Introducing a new framework for negotiating BDSM participation. *Electronic Journal of Human Sexuality*, 17.
<https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=15455556&v=2.1&it=r&id=GAL E%7CA419762806&sid=googleScholar&linkaccess=abs>
- Wolf, S. (1990). *Freedom Within Reason*. Oxford University Press.
- Wyatt, N. (2009). Failing to Do Things with Words. *Southwest Philosophy Review*, 25(1), 135–142. <https://doi.org/10.5840/swphilreview200925114>