

The Responsible Self in the Face of Situationist and Automaticity Challenges

Sue (Yifan) Su, *College of William and Mary*

Abstract: Contemporary situationist and automaticity research poses a great challenge to our traditional conception of moral responsibility and requires a picture of agency that accounts for these compelling empirical results. In light of the emerging scientific evidence, we might need to reevaluate our intuitions about the close connection between responsible agency and issues of choice and control. Appealing to Sripada's care-based conception of the real-self and motivational support account of self-expression, this paper aims to preserve moral judgment and responsibility, as well as a distinct sense of control. As an alternative to the rationalistic and identificationist accounts centered around conscious states and processes, this care-based view takes a relationist and pluralist approach to the questions "what constitute the real-self" and "how do actions reflect the real-self". Based on the wide-ranging functional properties and action-directed mechanisms, unreflective cares fundamental to an agent's identity incline channels outside of conscious awareness that issue in care-promoting actions, thus accounting for a broad array of conducts that are non-conscious or non-deliberative but nevertheless self-expressive. We might not have a conscious deliberative self with as much direct control as we expected, but we have a responsible self with ample control to mitigate the situationist and automaticity challenges.

Background

Contemporary situationist and automaticity research poses a great challenge to our traditional conception of moral responsibility, and requires a picture of agency that accounts for these compelling empirical results.¹ Situationist literature shows that various environmental features influence our actions without our awareness and even contrary to our intentions and expectations.² Meanwhile, automaticity literature suggests that conscious deliberation and control are often bypassed or overridden by automatic sub-cognitive processes.³ The source of many, if not the majority of, our day-to-day actions thus seems to lie in such situational factors and automatic processes—forces we have previously underestimated or neglected—that are separate or even disconnected

¹ Manuel Vargas, "Situationism and Moral Responsibility: Free Will in Fragments," in *Decomposing the Will*, ed. Tillman Vierkant et al. (New York: Oxford University Press, 2013), 325-349.

² Dana Nelkin, "Freedom, Responsibility, and the Challenge of Situationism," *Midwest Studies in Philosophy* 29, no. 1 (2005): 181-206, PhilPapers; Alfred Mele and Joshua Shepherd, "Situationism and Agency," *Journal of Practical Ethics* 1, no. 1 (2013): 62-83, PhilPapers.

³ John Bargh and Tanya Chartrand, "The Unbearable Automaticity of Being," *American Psychologist* 54, no. 7 (1999): 462-479, APA PsycNET; Michael Brownstein and Alex Madva, "The Normativity of Automaticity," *Mind and Language* 27, no. 4 (2012): 410-434, PhilPapers.

from our conscious deliberative self. In addition, there is growing empirical evidence—complementary to the situationist and automaticity research—of a “boundedly rational” mind driven by non-conscious associative processes and with limited intentional conscious control.⁴ All of these result in the apparent loss of control and undermining of agency, threatening our conceptions of freedom and responsibility for actions, and our intuitions about the close connection between responsible agency and issues of choice and control.

General Real-Self Views

It is intuitive that mental states or attitudes that are genuinely our own are the source of free and responsible actions, rendering the general real-self view an attractive theory. The notion of real-self marks out an agent’s distinct mental states that produce and govern his actions and are relevant to moral judgments as the locus of agency and responsibility. Amongst real-self theorists, the exact element with agential authority and the exact nature of an agent’s identification relation to that element are much disputed. However, identificationist views always require that an agent is consciously aware of and evaluates his desires and motives, and decisively identifies with what he wants to be effective or judges to be desirable, for those mental states to genuinely belong to him.⁵ Preliminary problems these views face include the regress of identification, and the assumption of a rational being with reflective capacities.

Advancing a less rationalistic account of the real-self, Sripada selects cares as the agential element.⁶ Cares are defined as a distinct class of pro-attitudes of importance and fundamental to oneself, which play a characteristic functional role in motivating actions. Distinct from a conscious deliberative self, a care-based responsible self could account for a broader array of non-deliberative but self-expressive conducts—including those faced with situationist and automaticity challenges—that pose a problem for the highly agentially-demanding views. An agent can be responsible for actions that are expressive or reflective (used interchangeably) of his cares that are subsets of the real-self, but are nevertheless influenced by situational features or unconscious processes. For one’s cares to constitute his real-self, and for his care-promoting action to reflect the real-self, an agent need not be conscious of his cares that motivate his action at the moment.

Conception of Self: Cares v. Conscious States

Is consciousness truly indispensable for mental states or attitudes to genuinely belong to the real-self as rationalists suggest? According to Sripada, inner sources including conscious reflective thoughts—despite lying within the boundaries of one’s psychology—

⁴ Michael Brownstein, “Attributionism and Moral Responsibility for Implicit Bias,” *Review of Philosophy and Psychology* 7, no. 4 (2016): 765-786, PhilPapers; Bargh and Chartrand, “The Unbearable Automaticity of Being”.

⁵ Harry Frankfurt, “Freedom of the Will and the Concept of a Person,” *Journal of Philosophy* 68, no. 1 (1971): 5-20, PhilPapers.

⁶ Chandra Sripada, “Self-Expression: A Deep Self Theory of Moral Responsibility,” *Philosophical Studies* 173, no. 5 (2016): 1203-1232, PhilPapers.

could be alienated from an agent's real-self.⁷ Reflective and endorsed attitudes or judgments might fail to articulate one's genuine perspective, or might be simply mistaken.⁸ Empirical evidence suggests that information encoded at the conscious level does not necessarily have greater accuracy compared to automatic evaluations.⁹ An example might more clearly illustrate this. Jack firmly believes in the importance of civility in the workplace and judges politeness to be a desirable quality. In moments of stress, however, he treats his co-worker Amy disrespectfully and focuses solely on his tasks at hand, without conscious endorsement of or identification with such behavior.

In this case, there is minimal voluntary control, explicit choice, or conscious beliefs associated with Jack's rude behavior towards Amy; meanwhile, he consciously endorses ideals of politeness and civility. Rationalists or identificationists would likely excuse Jack's rudeness. Levy, for example, would highlight his conscious beliefs in politeness as the real-self, discarding any non-conscious and unreflective mental states as not genuinely Jack's.¹⁰ However, it might be tempting to locate some anchoring of Jack's rudeness in his self—we might say "he (some part of him) was surprisingly rude to Amy"—rather than deny any connection to his real-self, and say "Jack is never rude" or "that rude guy is not Jack". It might also be tempting to assign judgment especially if Jack's rudeness marks a pattern over time—we might say "the rude Jack is terrible". Thus, it might be unreasonable to consider Jack's real-self as only constituted by his conscious evaluative judgments and beliefs. Sripada's notion of cares offers an alternative—certain non-rational psychological attitudes, cares, which might be unconscious, constitute Jack's real-self and motivate his unreflective rude actions. In stressful situations at least, meeting deadlines and performing his tasks well, possibly at the cost of civility, are of great importance to Jack. Such cares—more fundamental and integral to Jack's self—seem to bypass and contradict his conscious beliefs that are alienated from his real-self in the given situation.

Cares are inherently internal to an agent and underwrite one's identity as the source of deep or real-self.¹¹ They serve a characteristic functional role in inclining one's intrinsic motivation, practical reasoning, and emotional connection in favor of achieving care-promoting actions. As the source and foundation of motives for actions, cares shape positive "evaluative, commitmental, and emotional" effects in the prospect of care-promoting actions, exhibiting functional properties that are "conceptually tied" to what is of importance to the agent and fundamental to his identity, thus constituting the real-self.¹² Importantly, cares define the ontological conception of real-self, which is to be distinguished from the psychological one—what *in fact* constitute the real-self based on the characteristic functional roles does not depend on what one *takes* to be fundamental.¹³ Therefore, whether one consciously identifies with certain contents of the self, they

⁷ Ibid., 1204.

⁸ Ibid., 1207-8, 1212.

⁹ Bargh and Chartrand, "The Unbearable Automaticity of Being," 475.

¹⁰ Neil Levy, "Expressing Who We Are: Moral Responsibility and Awareness of Our Reasons for Action," *Analytic Philosophy* 52, no. 4 (2011): 243-261, PhilPapers; Neil Levy, "Consciousness, Implicit Attitudes, and Moral Responsibility," *Noûs* 48, no. 1 (2014): 21-40, PhilPapers.

¹¹ Brownstein, "Attributionism and Moral Responsibility for Implicit Bias," 773.

¹² Sripada, "Self-Expression," 1209-11.

¹³ Ibid., 1211-2.

constitute the real-self and serve as the basis for moral responsibility as long as they are cares characterized by the relevant properties.

If cares fulfill such essential functional roles that they constitute the real-self, why does “what we are conscious of” often seem closely and uniquely associated with “what is ours”? What might explain such a strong intuition is that consciousness of our attitudes conveys a sense of voluntary control and choice involved in owning or integrating those attitudes as “ours”. King and Carruthers highlight the important role of reflective reasoning and endorsement in either generating states that belong to one’s real-self, or *appropriating* into the self those states imposed upon oneself that originate from the outside.¹⁴ Levy similarly argues that consciousness of the content of attitudes is necessary for their *integration* into an agent’s self-conception.¹⁵ The irresistible desire due to addiction is an example of a mental state whose origin lies outside of the agent, and which is alien to the self unless appropriated by conscious reflection.

However, Jack’s cares do seem to have unifying effects in constituting his identity and his integrated disposition as a rude and accomplishment-driven person *in stressful situations*. On the other hand, his explicit conscious beliefs in politeness—which supposedly constitute a relatively coherent and person-level concern or perspective—seem to fall short of imposing unity on his identity in the given situation. What seems more integrated into his real-self is the cares that lie outside of his awareness and voluntary control. This strongly suggests that unreflective cares could achieve the kind of unification of a person on which moral agency arguably depends. Smith similarly argues that our moral practice is not an activity of voluntary choice or control; what is “ours” reflects what we judge to be of value, whether we consciously and voluntarily choose and endorse it.¹⁶

Expression of Self: Motivational Support v. Identificationism

For an action to express the real-self and thus be morally judged, rationalistic accounts of expression require conscious mental states that cause actions and processes to produce, sustain, and govern those states. For example, Frankfurt highlights reflective self-evaluation of one’s desires and motives—either endorsement or criticism—and eventually decisive commitment to the desire that one wants to be operative.¹⁷ Levy similarly argues that the degree of personal control and consciousness marks different ranges of agency and responsibility in actions,¹⁸ and that one needs to consciously identify with the “personally available” mental states—easily retrievable and poised to guide his behavior—in order for it to fully express one’s stance.¹⁹

Returning to our example of the rude Jack, where we previously have the temptation to attribute his rudeness to his real-self, it might also be reasonable to consider his action

¹⁴ Matt King and Peter Carruthers, “Moral Responsibility and Consciousness,” *Journal of Moral Philosophy* 9, no. 2 (2012): 219, PhilPapers.

¹⁵ Levy, “Consciousness, Implicit Attitudes, and Moral Responsibility,” 36.

¹⁶ Angela Smith, “Responsibility for Attitudes: Activity and Passivity in Mental Life,” *Ethics* 115, no. 2 (2005): 237, PhilPapers.

¹⁷ Frankfurt, “Freedom of the Will and the Concept of a Person,” 7-8.

¹⁸ Neil Levy and Tim Bayne, “Doing without Deliberation: Automatism, Automaticity, and Moral Accountability,” *International Review of Psychiatry* 16, no. 4 (2004): 213, PhilPapers.

¹⁹ Levy, “Expressing Who We Are,” 246.

as attributable to and reflective of at least some part, however small, of the real-self, rather than dismiss his action as out-of-character or disconnected from the self. Instead of simply taking a rationalistic approach that considers Jack's unreflective and unendorsed rude behavior as non-self-expressive, it might be best to explore Sripada's motivational support account of expression.

As an alternative to the highly reflective enterprise involved in Frankfurt's endorsement-based account, Sripada's account of the expression relation hinges on the motivational and causal influences of cares on attitudes and actions. Both Brownstein and Sripada highlight the causal power of cares in motivating actions that express something morally important about oneself.²⁰ For expression to occur, cares must exert influences—corresponding to their functional properties aforementioned including motivational, commitmental, evaluative, and emotional elements—on wide-ranging action-directed mechanisms that issue in care-promoting actions. Cares incline processes of evaluative deliberation, habitual reinforcement, and emotional appraisal—channels within or outside of conscious awareness that motivate actions.²¹ Thus, an agent does not necessarily *need* to comprehend or judge consciously his cares anchored in his real-self that motivate and govern actions, nor does his reflective judgment or deliberation *need* to align with his cares. Actions—including those that are non-conscious, non-volitional, or even divergent from one's will—are open to evaluation and an agent subject to moral judgment, since one's actions issued by aforementioned mechanisms do reflect upon his cares and express moral agency.²²

Why does consciousness seem essential for actions to be self-expressive and genuinely responsible? Indeed, the motivational support account faces difficulties, especially the seemingly obscure operation of action-directed mechanisms through which cares issue in actions. Meanwhile, the way consciousness provides a direct link between actions and the self—and thus factors into the expression relation—is straightforward and intuitive. Actions motivated by conscious attitudes seem to naturally express one's overall evaluative stance and the real-self. However, such rationalistic identificationist accounts operate on atomistic and monistic assumptions underlying our common understanding of free will.²³ An agent is responsible for his actions only when he identifies with some special agential element, a single unified cross-situationally stable entity, via monistic self-governing mechanisms involving only conscious endorsement or evaluative judgment—independent and irrespective of contextual inputs from past or present circumstances in response to which our attitudes might automatically move us to act.

However, various unconscious agential systems shown to be relevant and even pervasive in our daily actions, along with the limitations of our conscious mechanisms and capabilities, seem to challenge our traditional conceptions of agential elements and mechanisms. Given the empirical and conceptual evidence of automatic self-regulation processes,²⁴ spontaneous attitudes and reactions,²⁵ automatic tension-alleviating

²⁰ Brownstein, "Attributionism and Moral Responsibility for Implicit Bias," 773; Sripada, "Self-Expression," 1215-6.

²¹ Sripada, "Self-Expression," 1217-9.

²² Brownstein, "Attributionism and Moral Responsibility for Implicit Bias," 767; Smith, "Responsibility for Attitudes," 263.

²³ Vargas, "Situationism and Moral Responsibility," 333.

²⁴ Bargh and Chartrand, "The Unbearable Automaticity of Being".

²⁵ Smith, "Responsibility for Attitudes".

responses,²⁶ and unreflective reason-directed actions,²⁷ it seems that consciousness does not necessarily play a more direct role in producing genuine and responsible actions. Therefore, an alternative to rationalistic accounts of expression abandoning atomistic and monistic assumptions might be more attractive. The motivational support account takes exactly such a relationist and pluralist approach to the expression relation. The following section shows how cares—as context-specific subsets of the real-self—are less intrinsic and robust than conscious reflective judgments, and how cares—through action-directed mechanisms—issue in actions much more heterogeneous than rationalistic mechanisms of conscious identification.

Application: A Relationist & Pluralist Approach

Applying the care-based conception of self and the motivational support account of expression that reject the requirement of conscious identification, we are able to preserve moral responsibility in a broader array of conducts and accommodate the compelling situationist and automaticity research. The earlier example of Jack can be framed as a case that poses situationist and automaticity challenges. Jack's unreflective rudeness might be overwhelmingly influenced by deliberatively-irrelevant situational features, such as his boss' passing comments about the importance of his project, the presence of his stressed colleagues, or his unpleasant experience earlier in the day with his lunch order. Presented with such environmental stimuli, Jack's motives and actions might originate from automatic mechanisms that operate effortlessly and without conscious guidance—through an automatic *evaluation* of the importance of task performance, and the automatic *activation of goals* or cares for task performance and achievement, due to situational features or habitual and emotional effects.²⁸

Abandoning the assumption of atomism (or internalism), Sripada's mosaic conception of the real-self is relationist, allowing for conflict of cares and taking into account the situational influences. Jack's real-self can be understood as involving rich complexities and occasional conflicts—there exist distinct subsets as well as divergent sets of prioritization that exert causal and motivational influence on his actions in different contexts.²⁹ The source of his rude actions is not solely internalist and non-relational; past and present contexts factor into cares and thus self-reflective actions—the cue of stress in Jack's past and present situations is closely connected to his fundamental cares for achievement. Thus, the appearance that the environmental features cause such actions and that his rudeness is non-self-reflective can be explained as situational triggers and cues mapping onto subsets of Jack's real-self. His rude treatment of Amy is likely motivated by his cares for task performance in *moments of stress*. In a different context without the cue of stress, Jack might prioritize his belief in politeness and act civilly towards Amy.

The motivational support account is pluralist rather than monistic. Instead of serving as a single agential structure, cares motivate actions through a variety of distinct action-directed mechanisms, and thus a wide range of actions can be expressive of the real-self.

²⁶ Brownstein and Madva, "The Normativity of Automaticity".

²⁷ Caroline Arruda and Daniel Povinelli, "Chimps as Secret Agents," *Synthese* 193, no. 7 (2016): 2129-2158, PhilPapers.

²⁸ Bargh and Chartrand, "The Unbearable Automaticity of Being".

²⁹ Sripada, "Self-Expression," 1225-6.

Jack's cares for meeting deadlines and performing well can provide motivational support for actions through channels covering a wide spectrum of consciousness.³⁰ Via conscious assignment of evaluative weights, non-deliberative reinforcement learning, or automatic production of affective markers, cares incline his deliberative *or* automatic processes in favor of rude actions.³¹ However, his conscious evaluation might not and *need* not be involved, as his rude actions could be simply motivated through channels outside of his awareness, such as via spontaneous learning of habits and the automatic operation of his affective system—the prospect of rude treatment of Amy is likely affectively marked as positive based on past situations where Jack's cares are satisfied, reinforcing this tendency and pattern of rude actions.

Sense of Control Revisited

Despite the apparent lack of control in cases of situationist and automaticity research, as well as the limited capacity and role of conscious processes in general, this paper advances a positive view about moral responsibility. The care-based conception of self and motivational support account of expression help preserve not only moral responsibility, but also a sense of control—albeit distinct from the one commonly associated with a conscious deliberative self—which seems to be of intuitive importance to responsible agency. Specific issues of control are beyond the scope of this paper, but the following brief suggestions might shed light on different ways of understanding control.

Instead of being aware of our desires and motives at the moment of action, a different sense of control lies in forming causally efficacious implementation intentions to carry out plans that promote intentional goals.³² Jack might reflect on his cares after his rude treatment of Amy, and commit to improving his actions in the future in a similarly stressful situation. By means of such conscious distal intentions, Jack could mitigate the effects of his unreflective cares on his reaction time while promoting his conscious belief in politeness—an example of exercising control.³³ Through practice and habituation, Jack could exercise long-range control on his cares for achievement at the cost of civility in moments of stress—impossible to be willed away yet susceptible to revision—by means of indirect self-regulation strategies including evaluative conditioning and mental imagery.³⁴ Additionally, Jack might improve his moral vision—the capacity for perceiving morally relevant considerations—so as to detect situational effects of stress that map onto his cares, and to limit similar motivating forces in the future. Indeed, the conscious deliberative agency does play a role—albeit not direct or immediate as we thought—in resolving conflicts between unreflective attitudes and conscious intentions to regulate actions.³⁵ We might not have a conscious deliberative self with as much direct control as we expected, but we still have a responsible self with ample control to mitigate the situationist and automaticity challenges.

³⁰ *Ibid.*, 1216.

³¹ *Ibid.*, 1217-9.

³² Mele and Shepherd, "Situationism and Agency," 77.

³³ *Ibid.*, 74.

³⁴ Brownstein, "Attributionism and Moral Responsibility for Implicit Bias," 770, 781.

³⁵ Vargas, "Situationism and Moral Responsibility," 340-1.

References

- Arruda, Caroline T. and Povinelli, Daniel J. "Chimps as Secret Agents." *Synthese* 193, no. 7 (2016): 2129-2158. PhilPapers.
- Bargh, John A. and Chartrand, Tanya L. "The Unbearable Automaticity of Being." *American Psychologist* 54, no. 7 (1999): 462-479. APA PsycNET.
- Brownstein, Michael. "Attributionism and Moral Responsibility for Implicit Bias." *Review of Philosophy and Psychology* 7, no. 4 (2016): 765-786. PhilPapers.
- Brownstein, Michael and Madva, Alex. "The Normativity of Automaticity." *Mind and Language* 27, no. 4 (2012): 410-434. PhilPapers.
- Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68, no. 1 (1971): 5-20. PhilPapers.
- King, Matt and Carruthers, Peter. "Moral Responsibility and Consciousness." *Journal of Moral Philosophy* 9, no. 2 (2012): 200-228. PhilPapers.
- Levy, Neil. "Expressing Who We Are: Moral Responsibility and Awareness of Our Reasons for Action." *Analytic Philosophy* 52, no. 4 (2011): 243-261. PhilPapers.
- Levy, Neil. "Consciousness, Implicit Attitudes, and Moral Responsibility." *Noûs* 48, no. 1 (2014): 21-40. PhilPapers.
- Levy, Neil and Bayne, Tim. "Doing without Deliberation: Automatism, Automaticity, and Moral Accountability." *International Review of Psychiatry* 16, no. 4 (2004): 209-15. PhilPapers.
- Mele, Alfred R. and Shepherd, Joshua. "Situationism and Agency." *Journal of Practical Ethics* 1, no. 1 (2013): 62-83. PhilPapers.
- Nelkin, Dana K. "Freedom, Responsibility, and the Challenge of Situationism." *Midwest Studies in Philosophy* 29, no. 1 (2005): 181-206. PhilPapers.
- Smith, Angela M. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, no. 2 (2005): 236-271. PhilPapers.
- Sripada, Chandra. "Self-Expression: A Deep Self Theory of Moral Responsibility." *Philosophical Studies* 173, no. 5 (2016): 1203-1232. PhilPapers.
- Vargas, Manuel. "Situationism and Moral Responsibility: Free Will in Fragments." In *Decomposing the Will*, edited by Tillman Vierkant, Julian Kiverstein, and Andy Clark, 325-349. New York: Oxford University Press, 2013.