

INTEGRITY AND SPONTANEITY

Jonathan Gingerich

9 January 2021

Abstract. Many of us have experienced a peculiar feeling of openness and spontaneity, associated with the ideas that our lives could go in many different directions. Whilst this sort of spontaneity is widely celebrated in popular culture, I argue that such spontaneity conflicts with a popular account of the self. According to that account—endorsed by a number of moral philosophers but most famously associated with Christine Korsgaard—the ideal self is one that is well-ordered, actively constituted through self-conscious rational deliberation. Such *reflective integrity theories* of the self are attractive because they provide an account of agency according to which ideal moral agents invariably endorse their own actions and beliefs. But I argue that integrity theories forbid well-constituted agents to seek out experiences of spontaneity: such agents must act in a manner that they antecedently and reflectively endorse, whilst spontaneity requires acting in a manner not antecedently fixed by one’s own decisions. Insofar as we do and should value spontaneous activity, my account exerts pressure on reflective integrity theories of the self.

‘Caught—the bubble / in the spirit level, / a creature divided; / and the compass needle / wobbling and wavering, / undecided. / Freed—the broken / thermometer’s mercury / running away; / and the rainbow-bird / from the narrow bevel / of the empty mirror, / flying wherever / it feels like, gay!’ – Elizabeth Bishop, ‘Sonnet’ (Bishop 1979, 38)

§ 1. Introduction

What is the proper relationship between rational, deliberative capacities and the self? It is widely agreed that our abilities to reflect and deliberate about what to believe and do are central to our selfhood, but it also seems that we have desires, instincts, and feelings that are part of us but are distinct from these rational capacities. Furthermore, it seems that we can make ourselves into the sorts of people who are more fully defined by their integrity, coherence, and rationality, or else into the sorts of people who are defined to a greater degree by their impulses, instincts, or feelings. Ethically speaking, which sort of self should we aspire to be? Should we aspire to maximise the extent to which we are defined by our rational capacities, or should they play a more qualified role in our understanding of an ideal self?

Integrity theories answer these questions by contending that our rational, deliberative capacities must *integrate* the other parts of the self in order for us to be agents, in order for us to live up to our moral duties, or in order for us to count, metaphysically, as selves.¹ One version of integrity theory, *reflective* integrity theory, maintains that, for us to be selves or agents, or to be ethically successful selves or agents, we must consciously and reflectively guide and control who we are and what we do.² One of the leading contemporary proponents of reflective integrity theory, Christine Korsgaard, contends that we should, to the extent that we can, make ourselves into the sorts of people whose actions are guided by self-consciously and reflectively endorsed ‘practical identities’ that ‘necessitate’ our actions in a law-like way. Korsgaard’s theory of the self is compelling because it produces an account of agency according to which agents who successfully self-constitute are guaranteed to endorse their actions and beliefs and are guaranteed to commit themselves to the norms of Kantian morality. Insofar as we succeed at reflectively self-constituting, we are guaranteed to be in control of who we are and what we do, and we are certain not to be estranged from ourselves

Korsgaard’s theory of self-constitution captures a sort of selfhood that many people (and perhaps almost everyone) aspires to, at least in certain moments. However, such a theory comes with steep costs when it contends that we can be integrated as selves and agents to greater or lesser degrees and that we should seek to unify and integrate ourselves

¹ Versions of integrity theory, which vary a great deal, include Bratman 2007; Crowell 2007; Korsgaard 1996; 2000; 2008; 2009; Millgram 1997; and Velleman 2000; 2006; 2009. Many other theories of the self contend that the disparate parts of the self must be integrated in some manner without contending that the self’s rational capacities must *perform* this integrating function (see Strawson 2009); I do not include such theories under the heading of ‘integrity theory’.

² Proponents of this variety of integrity theory include David Velleman (2000; 2006; 2009) and Christine Korsgaard (1996; 2000; 2008; 2009). For a discussion of how the role of self-reflection differs in Velleman and Korsgaard’s constitutivist theories of the self, see Arruda 2016. In this essay, I focus on Korsgaard’s view, rather than Velleman’s, because it more clearly proposes that we should maximise the degree to which our actions are controlled by our rational, deliberative capacities than does Velleman (cf. Velleman 2009, 26-27).

to the maximum degree that we can. Particularly, I will argue that this demand precludes well-constituted agents from experiencing *spontaneity*, where the sort of spontaneity that I have in mind is the spontaneity that we speak of when we call someone a ‘spontaneous person’ or that we might imagine to play a central role in the life of a ‘free spirit’. To experience spontaneity requires, among other things, that one’s actions not be settled in advance, either by one’s own conscious, deliberative plans and decisions or those of other agents. Part of what makes spontaneity valuable is that it provides us with an experience of outstripping our rational, deliberative nature.

In §2 of this essay, I describe Korsgaard’s integrity theory and explain why its ethical view of the self is compelling. In §3, I describe the experience of spontaneity and defend its value before showing why Korsgaard’s theory rules out the possibility that an ethically well-constituted agent could aspire to experience spontaneity. Whilst Korsgaard’s theory accommodates a great deal of creativity and allows for certain restricted forms of spontaneity, it ultimately precludes some of the central features of spontaneity—features that are part of why spontaneity is worth wanting. Finally, in §4, I consider and respond to a transcendental argument that reflective integrity theorists might raise against the picture of spontaneity and the self that I present here.

§ 2. Integrity theory

In *Self-Constitution: Agency, Identity, and Integrity*, Korsgaard sets out to show that, to lead a good life, one must achieve a high level of integrity as an agent. A good life is one that is unified and whole. Living a life of integrity, for Korsgaard, involves engaging in the activity of self-constitution and succeeding in the ‘struggle for psychic unity, the struggle to be, in the face of psychic complexity, a single unified agent’ (Korsgaard 2009, 7). Being a single unified agent requires acting in such a way that one’s actions arise from ‘the person as a whole’ rather than issuing from ‘forces working in or on an agent’ (133-34). Actions that arise from the whole person stand apart from those that do not in their ‘necessitation’

by one of the practical identities of a unified agent, which ‘include such things as roles and relationships, citizenship, memberships in ethnic or religious groups, causes, vocations, professions, and offices’ (20). These practical identities all provide ‘absolute inviolable laws’ to guide an agent’s choices (23). For instance, one person might help another person out because she is his mother, and the practical identity of ‘mother’ provides the person with action-guiding principles along the lines of ‘help your children accomplish their goals when you can’ (21-22).³ On this view, a good action ‘is one that both achieves and springs from the integrity of the person who performs it’ (25).

Achieving integrity further requires agents to resolve conflicts among their practical identities when the principles provided by those identities come into conflict, as a failure to do so would be a failure to live up to the standards provided by one or more of one’s practical identities (Korsgaard 2009, 25). Agents cannot simply combine their various practical identities in any manner that they like. Integrity requires agents to will that the principles or laws generated by *each* of their practical identities apply to *all* similar cases of willing that the agent will encounter in the present or the future. This is because an agent who wills particularistically—treating a reason as applicable only to the case at hand—fails to act as a single agent at all (72-76).⁴ The impossibility of particularistic willing also

³ To achieve integrity, an agent must have at least one practical identity, because without *some* practical identity ‘you will lose your grip on yourself as having any reason to live and act at all’ (Korsgaard 1996, 121).

⁴ ‘Universal willing’ can be ‘provisionally universal’ for Korsgaard, which means that the generality of a universal reason may not be fully determined. If I will to get up as soon as my alarm clock goes off in the morning, I might satisfy the requirements of universal willing if, tomorrow morning when I hear my alarm go off, I hit the snooze button because I decide in that moment that I should not get up as soon as my alarm goes off when I have not slept for at least six hours. On the other hand, I would will particularistically if, when I heard my alarm go off, I hit the snooze button because I merely wanted (e.g., had an occurrent desire) to stay in bed a while longer. In this case, I would be willing for reasons that conflict with the reasons for which I willed my wake-up-with-the-alarm norm. (This example is meant only to illuminate the distinction between universal and particularistic willing; it might be obtuse, or even irrational, to will universally ‘to get up as soon as my alarm clock goes off,’ for such a willing is so granular and disconnected from my deeper interests that it would be bizarre to universally will this particular maxim.)

implies, on Korsgaard's telling, that 'the categorical imperative is a constitutive principle of action' (72) and that successful self-constitution requires a commitment to principles of Kantian morality (214).

Korsgaard argues that this account of what it is to be a good *agent* is also the correct account of what it is to be a good *human*. As human beings are 'reflective animals' who seek reasons that tell us what to do and how to live (Korsgaard 2009, 115-16), we are 'condemned to choice and action' (1). Insofar as we aim to have reasons for how we act and live, we are committed to seeking universally applicable reasons for how we can act and live as single and unified individuals. Any creature whose mind has 'reflective awareness of its mental state' is committed to seeking integrity, seeking to unify themselves by living up to their own standards (15-16), so being a good agent is the only thing that is choiceworthy for a person (177). Thus, for Korsgaard, living a successful human life consists in being a successful agent, being an agent requires being unified as a single thing, and achieving unity consists in achieving integrity in self-constitution through action.

Insofar as Korsgaard aims to explain what, metaphysically speaking, an agent *is*, it might seem at first glance that it is not possible to act badly, or to be a defective agent or self on her view: either you succeed at self-constitution and so you are a single, unified agent, or else you are a 'mere heap of unrelated impulses' (Korsgaard 2009, 76). But Korsgaard's theory of integrity is a practical ideal for agency, one that we may never live up to fully. When we face threats to our integrity, our challenge is to meet them well, and we can do a better or worse job of unifying ourselves as agents through action (163). Defective action occurs when we get close enough to self-constitution for our activity to qualify as 'action', but where this activity is 'badly done' (132). Defective action is *defective*, rather than bad according to some standard external to action, because it falls sort of the 'constitutive standards' that apply to it in virtue of the kind of thing that it is (28). For action, such a standard is provided by the aim of integrating and constituting oneself

through action. Autonomous action is ‘*defective* as autonomous action’ for Korsgaard when we succeed to some degree at constituting ourselves but when we do not fully ‘pull [ourselves] together’ in a principled way by making laws for ourselves (213-14). When this happens, our actions and selves are defective in the same way that a house with a leaky roof is defective: by failing to live up to the standards that are intrinsic to the ideas of ‘action’, ‘self’, and ‘house’, respectively (32).

To understand how an activity might be sufficiently in the ballpark of ‘action’ for the constitutive standards of action to apply to it whilst still falling well short of what those standards demand, it will help to consider Korsgaard’s story of ‘Jeremy’, her example of a ‘democratic soul’:

Jeremy, a college student, settles down at his desk one evening to study for an examination. Finding himself a little too restless to concentrate, he decides to take a walk in the fresh air first. His walk takes him past a nearby bookstore, where the sight of an enticing title draws him in to look at the book. Before he finds it, however, he meets his friend Neil, who invites him to join some of the other kids at the bar next door for a beer. Jeremy decides to have just one, and he goes with Neil to the bar. While waiting for his beer, however, he finds that the loud noise in the bar gives him a headache, and he decides to return home without having the beer. He is now, however, in too much pain to study. So Jeremy doesn’t study for his examination, hardly gets a walk, doesn’t buy a book, and doesn’t drink his beer (Korsgaard 2009, 169).⁵

Notably, Jeremy would be no better off, qua agent, if he *had not* gotten a headache at the bar and had finished his beer with Neil, because, in Korsgaard’s view, he would simply have lucked into drinking his beer. It is only an accident, Korsgaard thinks, that each of Jeremy’s impulses leads him to an action that undercuts the previous one, but it would also only be an accident if his impulses did not lead him to undercut his actions. Jeremy is, on Korsgaard’s constitutional model, ‘almost completely incapable of effective action’

⁵ Michael Bratman develops a very similar example, although he allows that a character like Jeremy might at least ‘accomplish a little bit with respect to each of several incompatible projects as [he] brute-shuffles from one to another’ (Bratman 2012, 83).

(Korsgaard 2009, 169) because it is just a matter of luck that he might end up doing anything at all. There is no constitutional principle in Jeremy that guides which desire he will act on. For Korsgaard, one can only succeed as an agent by acting *non-contingently* whilst Jeremy ‘is at the mercy of accident’ (169).

Jeremy’s activity includes enough deliberation and self-governance to count as action: he makes choices (e.g., to quit studying, to pop into the bookstore, to tag along with Neil) with some minimal degree of deliberation. What he lacks is some strong and principled way of draw his activity together into self-constituting action. He simply *finds himself* going for a stroll, in the bookstore, in the bar, and with a headache, rather than *making himself* through action. Korsgaard contends that we should aspire not to be like Jeremy, but instead to reflectively and deliberatively give ourselves laws that necessitate our actions.

The appeal of Korsgaard’s reflective integrity theory of the self is that when we fully succeed at self-constitution—when we do not allow ourselves to be governed by inclinations or laws outside of ourselves but instead govern ourselves—then we are guaranteed to be ourselves. People who are well-constituted according to Korsgaard’s requirements are guaranteed not to be alienated from themselves, insofar as self-alienation involves the sense that one’s life is not their own or that one is unable to move freely in their own life (Jaeggi 2014, 128). Any agent who successfully self-constitutes sees all of their life as their own because they reflectively endorse those practical identities that guide their actions and jettison all prospective actions that cannot fit with their identity. If I am fully self-governing, I will never, for example, just find myself screaming at my kids even though I hate the fact that I do this.⁶ Of course, Korsgaard does not think that I am likely to always succeed at pulling myself together; this is a normative ideal of action that is challenging to

⁶ I am grateful to an anonymous referee for suggesting this example.

live up to. And even fully constituting myself through my reflectively endorsed practical identities does not guarantee that I will never scream at my kids—it is at least conceivable that I or the world might change in such a way that I come to have reason to scream at my kids, in spite of my desire to avoid it. Nor does successful self-constitution guarantee that I will be happy, or that things will go well in my life. What successful self-constitution *does* guarantee is that I always be *myself*, made by myself, rather than by forces acting in or on me. As Korsgaard puts it, the self-governing, aristocratic person ‘is completely self-possessed: not necessarily happy on the rack, but herself on the rack, herself even there.’ (Korsgaard 2009, 180). Such a person might have occasion to unhappily scream at her kids, but she will never just *find herself* doing so. Korsgaard’s ideal of agency is attractive because, in striving to live up to it, we strive to guarantee that we are always *ourselves*.

§ 3. Spontaneity

Korsgaard’s theory of self-constitution has attracted a significant degree of scepticism from other philosophers of agency and the self, many of whom are unpersuaded by Korsgaard’s attempt to draw the objective validity of Kantian morality out of the nature of agency (see, *e.g.*, Pauer-Studer 2018; Velleman 2009, 147 n.35). Setting aside Korsgaard’s ambitions to ground morality, Korsgaard’s theory has an intuitive attraction: we might aspire to be fully self-governing because, in doing so, we save ourselves from being at the mercy of accident, at least when it comes to who we are and what we do. However, as I will argue in this section, selves that satisfy Korsgaard’s demands for successful agency cannot experience a valuable form of spontaneity, because the integrity and self-understanding achieved by Korsgaardian agents requires that their actions be ascribable to practical identities that they reflectively endorse, whilst spontaneity requires the *absence* of such deliberative endorsement.

My argument proceeds in two parts: first, I describe spontaneous activity and argue that it is a paradigmatically valuable human activity. Second, I argue that spontaneous

activity cannot be pursued by someone who aims to be fully self-governing in the sense required by Korsgaard's reflective integrity theory. I then consider and respond to several of the ways in which Korsgaard's reflective integrity theory might endeavour to make room for spontaneity as part of its ethical ideal of agency.

3.1 The nature and value of spontaneity

To get a sense of the nature of spontaneity, let us consider a modified version of Korsgaard's story about Jeremy.

Jamie, who identifies as a diligent university student, settles down in the library one evening to study for an accounting exam. Finding themselves stuck and feeling a bit deadened, they are a little too restless to concentrate. So, Jamie goes out for a little walk, not knowing where they will wander. As they stroll around the neighbourhood, Jamie pops into a bookshop to see if they have any study guides for accounting but gets distracted by an enticing volume of Blake's poetry. As Jamie is reading, their friend Ned comes across them and strikes up a conversation. Jamie and Ned get to talking about Blake, and Ned invites Jamie to come along to see a movie that is playing nearby. Caught up in the conversation with Ned and not really thinking about the accounting exam, Jamie puts down the volume of Blake and happily accompanies Ned to the theatre. At the end of the movie, Jamie observes to Ned that one of the characters seems a bit 'dead inside', and this observation sparks a conversation between Jamie and Ned about what it means to feel dead, whether the character really feels that way, and this says about Jamie's life. The conversation continues over one drink, and then another at a bar, until Jamie eventually heads back to their apartment, a bit too tipsy to study. Feeling slightly hung over the next morning, Jamie bombs the exam, changes their major to creative writing, and starts dating Ned. So Jamie doesn't study for their examination, gets a short walk, doesn't buy a book, sees a movie, changes majors, falls in love with a friend, and ends up feeling that they are 'a different person' than they were before.

Jamie's activity is *unplanned*, in that Jamie does not reflect or deliberate about and endorse the choices that shape their evening; they simply *find* themselves in a bookstore and then in a theatre and then at a bar. It is also *unscripted*, in the sense that it is not 'fixed by some preexisting "script", as with habit or rote rule-following as on an assembly line' (Ridge 2019, 4). Jamie's activity is *non-obligated*, in the weak sense that Jamie does not experience the course of their evening as dictated by their explicit or implicit beliefs about what other people, or the law, or morality demand of them, and it is *unalienated*, in the weak sense that

Jamie does not feel that anything that they do over the course of the evening is an alien imposition on them. The weak sense of non-obligatoriness and non-alienation suffice for Jamie to take pleasure in their activities. Spontaneous activity is, then, unplanned, unscripted activity that is phenomenologically non-obligatory and unalienated and that is typically, if not necessarily, pleasurable (see Gingerich n.d.). Such activity outstrips a self's conscious identity in the sense that the grounds of the spontaneous activity are not found in the self's conscious, first-personal reflection and endorsement.

As with Jeremy, it is 'only an accident' that each of Jamie's impulses leads them to an action that builds on 'the satisfaction of the last one' (Korsgaard 2009, 169). If things had gone slightly differently for Jamie, they might have found themselves in the same position as Jeremy, unable to accomplish anything with their evening. Like Jeremy, Jamie is 'at the mercy of accident' and just as 'the democratic life does not have to be like' Jeremy's evening (169), it also does not have to be like Jamie's. Thus, like Korsgaard's story of Jeremy, my story of Jamie is polemical, but much as the story of Jeremy draws out some of the dangers of living as a democratic soul, the story of Jamie draws out some of the values that might be realised by living a democratic life but not by an aristocratic one.

Jamie might point out that their spontaneity is different from 'living at random,' both because they do not feel alienated from their spontaneity (even if they do not reflectively endorse it) and because their spontaneity does not involve acting completely at random. Rather, they spontaneously choose among different practical identities that they care about as the evening progresses. What Jamie does over the course of the evening is, in every meaningful psychological sense, *up to them*, but at each juncture that they face, Jamie just follows their instincts rather than 'putting thought' into what to do. Whilst such a character's action is, according to Korsgaard, 'defective as autonomous action' (Korsgaard 2009, 162), it is Jamie's unique material and social circumstances and psychological states that produce spontaneous activity, not the role of a die. As Korsgaard could point out, this

still makes the Jamie's success *contingent*. But such contingent success suffices for the realisation of several values in Jamie's spontaneity.

Part of the value of Jamie's spontaneity comes from the good things that Jamie lucks into as a result of not taking their practical identity as a diligent student too seriously. We can imagine a version of Jamie—call them *Jamie the Good Student*—who fully endorsed this practical identity and used it to necessitate their actions in a law-like manner. Jamie the Good Student would place limit on how spontaneous their stroll could be to make sure that they made it back to the library to study. By not doing this, the Jamie of our story is in a position to 'luck in' to a variety of fun (talking with Ned over drinks after the movie), painful (the hangover), and exciting (switching to a new major) experiences that they would not have had these worthwhile experiences if they had governed themselves through a practical identity that they already endorsed. But these experiences are only part of the value of Jamie's spontaneity, and it is largely accidental that they comes about spontaneously—going for a stroll, reading Blake, and dating Ned are not obviously any more valuable because they arise spontaneously than they would be if they were planned or reflectively endorsed in advance.

But some of the value of Jamie's evening is less accidentally connected with Jamie's spontaneity. Jamie's spontaneity is *uncabined* in the sense that it is not authorized by any of Jamie's principles or practical identities that Jamie consciously endorses when they undertake the activity (although they may retrospectively embrace the activity).⁷ It might even be the case, if you had asked Jamie in the afternoon that they were studying if they would endorse a version of their life in which they failed their exam, changed majors, and

⁷ In contrast to uncabined spontaneity, we might think of *cabined* spontaneity as spontaneity that arises when agents 'voluntarily and temporarily give up agential unity in specific, controlled, and highly localized ways' (Nguyen 2019, 444-45) where reflectively endorsed practical identities provide a constitutional framework that delimits the scope of spontaneity.

started dating Ned, that they would not, although later on, having dated Ned for six months, Jamie would be happy to endorse what happened on the evening when they could not bring themselves to study for the accounting exam.⁸ This uncabinedness sets Jamie's experience apart from that of someone who *plans* for the sort of evening that Jamie has, or at least plans for its possibility, setting out from the library in hope of finding 'a little adventure'. Such a character, who we might call *Integrated Jamie*, could achieve the values that are ruled out by Jamie the Good Student's self-governance.

However, even Integrated Jamie would miss out on two important aspects of the value of Jamie's spontaneous evening. First, uncabined spontaneity provides Jamie with a feeling that their life encompasses a broader range of possibilities than they could plan for or deliberate about. Particularly, insofar as Jamie ends up with different practical identities and feels like they are a 'different person', Jamie might justifiably feel that they have an ability to surprise themselves that goes beyond their rational, deliberative capacities. Second, uncabined spontaneity also provides Jamie with a more passive experience of seeing that they are more than their reflective capacities. In allowing themselves to be pulled along by inclinations that have not come up for review in self-conscious reflection, Jamie can see their doings as products of nature flowing through them rather than of their own principled identification with a set of coherently organized practical identities. The value of seeing oneself as outstripping ones' rational capacities is available to agents who have experiences of uncabined spontaneity, but not to agents who experience spontaneity only when it is under the control of reflective deliberation.

⁸ As Benjamin Bagley has pointed out, we often 'simply find ourselves responding to concrete situations in particular ways, and we specify our values only incrementally and in retrospect, in order to make sense of our conduct as embodying them' (Bagley 2017, 870).

3.2 *The conflict between spontaneity and integrity*

Having presented the story of Jamie to demonstrate what spontaneous activity is and why it is valuable, I am now prepared to show how some of the most paradigmatically valuable experiences of spontaneity cannot be had by selves who satisfy the demands of Korsgaard's integrity theory. I will argue that valuable experiences of spontaneity make one seem to oneself to be metaphysically unstable because, at least from the first-personal perspective, it is not clear where one starts and where one stops: one acts from impulses or motivations that one takes to be non-identical with one's conscious, reflective standpoint and with which one does not antecedently identify

Korsgaard's view requires that practical identities provide us with laws that guide our actions, and laws 'lay[] down what is to be done' (Korsgaard 2009, 16). Succeeding as an agent involves an experience of 'necessitation' produced by 'work and effort' (7). Self-constitution happens at the level of conscious, reflective thought, and any actions that satisfy Korsgaard's requirements for agency must be intelligible as the products of practical identities' law-like necessitation. Further, leading a successful human life requires making oneself into an agent who acts in this manner rather than into someone who is divided against themselves. Anyone who satisfies this requirement cannot feel that what will happen with their life is *not* settled by their conscious, reflectively endorsed practical identities or by previously made decisions that are transparently available for conscious reflection.

Korsgaard might wish to regard 'spontaneous person' or 'free spirit' as practical identities that a successful agent might integrate into their other practical identities, but I will argue that Korsgaard cannot plausibly do so; her theory suggests that something is defective in any character who exhibits spontaneity of the sort that Jamie does. For Korsgaard, spontaneity must be suitably cabined so as not to give rise to intra-agential conflict. On Korsgaard's theory, 'if any possible change in my motivational state would

count as a good reason to do something other than what I am doing, then I am not making a decision, but merely observing the workings of the motivational forces within me' (Korsgaard 2009, 79). Yet this is precisely happens when Jamie experiences on the evening when they wander out of the library when they simply follow whatever impulses happen to occur: over the course of the evening, Jamie understands their agential role as one of observing and recording motivational forces that arise in them without their active participation.

Korsgaard might *partially* accommodate Jamie's spontaneity. Perhaps Jamie's spontaneous activity can be described at a higher level of abstraction, such what what is 'to be done' is 'live spontaneously'. Jamie might study in the afternoon because it 'feels like the right thing to do' at the time, but then take up Ned's proposal to go to the movie because, when it comes around, *it* feels like the right thing to do. If this were Jamie's practical identity, they would not be changing their mind without a good reason. Rather, they would act spontaneously in the afternoon and then again in the evening.

However, if this were Jamie's psychological makeup, Jamie would lack a Korsgaardian practical identity, because the maxim 'live spontaneously' has been constructed precisely to *avoid* specifying what is to be done by the agent. By reflectively endorsing the principle 'I will follow my inclinations where they lead me', Jamie would be 'heteronymous' rather than autonomous in that he would get 'his law from nature' by 'allow[ing] himself to be governed without much thought by its proposals' (Korsgaard 2009, 163). For this reason, agent who aspires to aristocratic self-government can aspire to have the spontaneous experiences of a democratic soul like Jamie. On Korsgaard's view, being a *good* human is incompatible with spontaneity insofar spontaneity is incompatible with successful agency, which all humans are committed to pursuing as far as possible.

Korsgaard's integrity theory is incompatible with any sort of agency that seeks the values realised by uncabined spontaneity.⁹

3.3 A Korsgaardian reply

Korsgaard might agree that stringent requirements for agential integrity rule out the possibility of experiencing uncabined spontaneity, but might argue that integrity still provides plenty of room for creativity and spontaneity and that whatever is *valuable* about spontaneity is provided for just as well by the forms of creativity and spontaneity that they allow. For instance, Korsgaard acknowledges that acts might be undertaken solely for their own sakes, as one might 'choose to dance for the sheer joy of dancing' (Korsgaard 2009, 12). Korsgaard also provides space for spontaneity in interpretation, argument, and creativity in determining what any practical identity requires. One might argue 'about whether a particular way of acting is the best way or the only way to go about being, say, a teacher or a citizen' and 'one might find a new way of being a friend' (21). So, for example, to the extent that I identify as an artist, I might exhibit creativity in working through what it means to be an artist in a particular context whilst my 'working through' is guided by my practical identity of 'artist'.

Whilst the forms of creativity that are accommodated by Korsgaard's integrity theory are important and valuable, they do not allow for some of the values centrally associated with spontaneity to be realised. Particularly, Korsgaard's integrity theory cannot allow Jamie's spontaneity to interfere with their other practical identities. Imagine an artist

⁹ There are less stringent theories of the integrated self that do not take principled action to require that one's choices be determined or narrowly constrained by action-guiding laws endorsed by conscious reflection. Ronald Dworkin, for instance, develops an account of integrity that allows for unreflective interpretation of one's values (Dworkin 2011, 101). Steven Crowell develops a Heideggerian theory of reflection of the sort that might integrate an agent that does not require explicit deliberation (Crowell 2007, 321). These views take complex, principled action to reflect an individual human's depth in a way that could not be contained in a list of rules or in any linguistic formulation. The 'principledness' of an agent's actions, on these views, reflects a commitment to working out the conflicts and lacunae that complex agents discover, over time, in themselves.

who has both a practical identity as an artist that involves spontaneous creativity and an identity as a punctual friend who does not show up late for dinner parties, Such an artist's spontaneity cannot be realised in a manner that interferes with their ability to show up on time for a dinner party (and vice versa). The spontaneity that is part of their identity as a creative artist has only as much free play as is permitted by all the artist's other practical identities. Adopting an identity that authorises spontaneity is compatible with successful self-constitution only if the spontaneity is suitably cabined. On Korsgaard's view, a well-constituted agent's deliberative, consciously endorsed practical identities must authorise all of their actions; thus Korsgaard acknowledges that we can '*choose* to dance for the sheer joy of dancing' rather than that we can *dance* for the joy of dancing (Korsgaard 2009, 12, emphasis added).

Korsgaard's accommodation of spontaneity fails to account for much of the reason that we might wish for spontaneity. Many people seek experiences of spontaneity in order to temporarily turn off or escape from these rational control systems that regulate their actions. Of course, there are risks to turning off these control systems: when you act spontaneously, you will find yourself doing things that you have not chosen, and you might find yourself doing things that you would not have chosen, had you deliberated. Elevating one's spontaneous dispositions above one's law-giving practical identities would involve failing to live up to the roles given by one's other practical identities because one could only 'luck in' to doing what their practical identities required.

Perhaps Korsgaard could also concede that it is ethically costly to rule out, as good forms of life, ways of living that involve uncabined spontaneity whilst arguing that the values achieved through spontaneity are non-moral values that are of lesser importance than the moral values associated with successful agency. Korsgaard might argue that the *point* of reflective self-constitution 'is liberation from the control of instinct' (Korsgaard 2009, 116). The aspiration to full self-governance is an aspiration to regulate our instincts so that,

whilst they ‘still operate within us, in the sense that they are the sources of many of our incentives’, they ‘no longer determine how we respond to those incentives, what we do in the face of them’ (116). Spontaneity, on this view, may be valuable so long as it is confined to things that aren’t all that serious—consumer choices about which brand of backpack to purchase, perhaps. Spontaneity can be part of human life but only so long as it is authorized and confined by a robust practice of choosing for reasons what to do. Non-agential values cannot matter for evaluating whether someone is a good human on this view because being a good human just is being a good agent and being a good agent is incompatible with uncabined spontaneity.

In my view, however, the experience of spontaneity is not just one valuable experience among many, but is a paradigmatically important form of human experience. Part of what it is to be human is not just to act based on reason, but also to have the experience of reason running out. And this is an experience that we can have only insofar as we are rational agents, insofar as we know what it is like to be governed by reason so that we can notice the difference when reason runs out.¹⁰

§ 4. A transcendental argument

A further argument remains in support of Korsgaard’s stringent version of integrity theory: an argument that by philosophising about what successful selves are or what successful agency is, we are already committed to the view that selves must satisfy the demands of integrity. In the remainder of this essay, I respond to this argument. I first describe the argument and then argue that it fails by demonstrating that reflective integrity theorists like Korsgaard are committed to regarding even people who fail at self-constitution as intelligible agents in some respects. However, to decisively reject Korsgaard’s transcendental argument, we must adopt a different understanding of the

¹⁰ I am grateful to Jordan MacKenzie for suggesting this line of thought.

nature of the self than that advanced by Korsgaard. A self can experience spontaneity provided that it includes a first-personal perspective on the world, some form of psychological and physical continuity, and psychological capacities that enable it to see itself in sources of action that are distinct from its rational nature.

Korsgaard argues in favour of her theory of self-constitution that by seeking justifications for your actions you commit yourself to being a single 'you', which you can only accomplish by having at least one practical identity and living up to all of your practical identities (Korsgaard 2009, 1) Otherwise, you are like a disunited city: not one thing, but many. I will call this Korsgaard's 'transcendental argument' for a reflective integrity theory of self-constitution.¹¹ This argument does not aim to show that it is 'desirable' or 'worthwhile' to be an integrated agent, but that you are already committed to being an integrated, unified agent, and it is on pain of inconsistency, of a sort, that you must acknowledge that being a good human requires being a good, integrated rational agent. You are a creature 'who needs reasons to act and to live.... [I]f you live at random, without integrity or principle, then you will lose your grip on yourself as one who has any reason to live and to act at all' (Korsgaard 1996, 121). If you want to be more than a mere heap, you have to decide how your identities cohere rather than just flitting from one to another *and* you must be committed to doing so as a single, unified agent, because the very fact that you are interested in having a philosophical conversation about the right way to live shows that you *care* about *what is choiceworthy*. '[B]eing human we must endorse our impulses before we can act on them' (Korsgaard 1996, 122). Because humans need practical conceptions of their own identities to find their actions worth undertaking, humans must

¹¹ I call this Korsgaard's transcendental argument because of its form. Korsgaard argues that the integrity norms of self-constitution are a necessary condition for the possibility of seeking justifications and, given that we do seek justifications, the truth of Korsgaard's integrity norms should follow.

take their practical identities to be 'normative,' such that these identities rationally necessitate their actions. 'If you had no normative conception of your identity, you could have no reasons for action, and because your consciousness is reflective, you could then not act at all. Since you cannot act without reasons and your humanity is the source of your reasons, you must value your own humanity if you are to act at all' (Korsgaard 1996, 123). If you worry about subordinating your attachment to some practical identities that you care about deeply, there must be a 'you' who is attached to those identities, and there can only be a single 'you' if you achieve integrity through self-constitution.

My first reply to this argument is to point out that experiences of spontaneity require (temporarily) *turning off* the intellectual drive that seeks reasons to act and to live. Experiences of spontaneity involve seeing yourself as unfixed by your deliberative decisions and consciously endorsed practical identities: they involve seeing yourself as acting in a way that reflects things outside of your rational self. Many practices that people rely on to experience the freedom from rational, deliberative control of their actions that characterises spontaneity make use of the connection between conscious self-awareness and the physical body. Consider two examples.

First, a tradition of Buddhist practice attempts to go beyond the experience of 'self' or 'I' as discrete and self-contained through a practice of meditation in which 'the activity of an individual practitioner's ego-consciousness' is arrested (Nagatomo 2017, §6.3). This experience is brought about, in substantial part, through an adjustment of body (including diet and exercise) and an adjustment of how one breathes. Practices of breathing, combined with the appropriate preparation of the body and the mind, give rise to an experience of 'kicking through the bottom of a bucket' and experiencing the self as 'a groundless ground that is nothing' (§8.2). Action then 'carries a sense of *spontaneity*' that surges 'from the creative source in the bottomless ground' (§8.2).

Second, psychiatrists have increasingly become interested in using hallucinogens to treat depression and existential anxiety in patients with terminal illnesses (Pollan 2018, 8-11). The experiences of users of psilocybin often involve losing track of the boundaries between self and non-self, feeling that they become other people (Grob 2007, 211). Subjects exposed to psilocybin often experience a temporary failure of the impulse for individuation—the breaking up of the world into separate, discrete persons and things. When this experience takes place in the context of supportive psychotherapy or nurturing relationships, subjects often connect this experience of the failure of individuation to their sense of self after the hallucinogenic session concludes and come to see themselves as less fully identified with their own rational impulses (Grob 2007). Such subjects often come to see death as less of a bad, because they feel less distinct from other people than they previously appeared to themselves (Grob 2007, 211).

The breathing practices of meditation and the use of hallucinogens are designed precisely to (temporarily) prevent one from asking questions like, ‘why should I have the aim of making sense?’ Absorbing oneself in an engrossing conversation, like Jamie’s conversation with Ned in the book shop, can have a similar effect. Temporarily severing the connection between self-awareness and theoretical intelligence often allows experiences of spontaneity to arise. It is not a puzzle how I can choose to meditate or consume a hallucinogen with the expectation that doing so will lead me not to ask questions like ‘why should I have the aim of making sense?’ for many people in fact do so. Korsgaard takes her transcendental argument to show that since rational action exists, it is possible, and since rational action is possible only if humans find their humanity to be valuable, we human beings must be valuable (Korsgaard 1996, 123-24). But actions that limit the scope of rational action (without eliminating the possibility of it) also exist. When people take such actions, they use their rational natures to choose (and value) something that is distinct from, and in tension with, their rational natures.

A reply to my counter-argument is available to Korsgaard. Claiming that people take *actions* to temporarily undermine their own rational agency or their tendency to act from a normative conception of their own humanity depends on the existence of some unified agent to whom such actions can be attributed. How can such an agent exist, other than through self-constitution? A reflective integrity theorist might say to a character like Jeremy or Jamie: ‘without reflectively deciding on how to act and who you are, there is no *you!*’

We might imagine Jeremy replying to this argument by saying: ‘I don’t have to be an “integrated self” to care both about studying and getting drinks with Neil. You told a coherent story about me trying to do both of those things! Too bad neither of them worked out, but maybe I’ll have better luck next time.’ The facts that Korsgaard’s story about Jeremy makes sense, that most of us, in fact, know (or else are!) ‘flaky’ people like Jeremy, and that sometimes people even try to make themselves *more* like Jeremy or Jamie (e.g., to be more ‘laid back’ or to ‘go with the flow’ or ‘chill out’) suggest that Korsgaard’s transcendental argument tries to accomplish too much. Even if Jeremy cares about what is (rationally) choiceworthy, he might also care about other values not grounded in rationality.

Korsgaard might reply to Jeremy’s rejoinder by asking how, unless he succeeds at self-constitution, he could know that he will even be the same person ‘next time’ he tries to study or go for a walk. But Jeremy can point out that he has lots of resources (memory, a name, a driver’s license, a body that changes only gradually over time) that allow him to be *somewhat* unified and that make him an intelligible conversational partner even if there is some other rational respect in which he is a ‘mere heap.’¹² Korsgaard asks rhetorically,

¹² For an account of a self that encompasses more than its conscious, deliberative nature, see Hubert Dreyfus’s reading of Heidegger on intentionality (Dreyfus 1993).

‘How do you interact with someone who is seriously divided against himself? *If you approach [a disunited city] as one city*, Plato says, *you’ll be making a big mistake*’ (Korsgaard 2009, 185, emphasis in original). A city constituted by interest groups that have simply made pragmatic alliances with one another is not best approached as a *city* but instead as a composition of potentially conflicting interest groups that, under the right circumstances, can be torn apart. Korsgaard suggests that you also make a big mistake if you approach a divided person as a person.

But even if Plato is right about treating a divided city that you hope to conquer as something other than a city, we might wonder: what is so difficult about talking to a person who both wants to stay out for another drink and also wants to get up at 6:00 tomorrow morning, although those desires conflict? Unlike the disunited city, you cannot talk to the two factions of the divided self separately from one another. The disunited self still has a single brain, only one mouth, and cannot be spatially separated into its constituent factions.

In defending his selfhood in this exchange, Jeremy must rely on a different notion of what a self is than does Korsgaard, regarding the self as something less rational and more embodied. Jeremy might, for example, regard the self as consisting in a single, mental subject occupying a first-personal viewpoint that is psychologically connected to a single physiological body that changes only gradually over time. This picture of the embodied spontaneous self need not claim that the self essentially *is* its body, or that its body cannot change. Some feature of a physical body—a voice, a prosthetic limb—might move from not counting as part of the self’s body to counting as part of the body as the self’s perspective on the body to which it is connected changes (Elliott 2003, 1-27). The embodied spontaneous self’s connection to a single, more or less temporally persistent body enables it to have attitudes toward its own future experiences that regards those experiences as *experiences of* the same self that it is now. Unlike successfully self-constituting Korsgaardian selves, such an embodied spontaneous self is not rationally guaranteed to be

the same entity in the future that it is now: its remaining a single entity depends on the contingent continuity of the physiological body that it is connected to. But this is all that the embodied spontaneous self needs to experience spontaneity.

It might appear that the conflict between Korsgaard's account of defective agency and her transcendental argument could be resolved by jettisoning her account of defective agency. However, this would be a substantial cost for integrity theory. Moreover, it would undermine the overarching project of integrity theory: to provide an account on which people are properly identified with their consciously endorsed rational plans and identities. The appeal of Korsgaard's transcendental argument depends on accepting the premise that the essence of human nature is rationality. If one thinks of rational agency as the essence of humanity or as an ahistorical feature of the experience of subjectivity, one may be drawn to the view that failures of the completeness of rational agency are failures of one's humanity. Whenever a person acts on 'a principle of choice which is not reason's own' the soul's unity is 'contingent and unstable' (Korsgaard 2009, 175). Korsgaard's integrity theory tells us that we should seek to overcome such contingency, so that our unity and stability is stable and guaranteed.

But if one thinks that human capabilities, including those capabilities constitutive of agency, have a biological or evolutionary history, and so that there was some point of time in the past at which agential capabilities developed, then one is likely to think that the stability that *any* human soul could have, even one governed by 'reason's own principle,' is contingent and unstable. From such a historical perspective, for humans to occasionally inhabit spontaneous practical identities and so to make the non-spontaneous components of their identity 'contingent' is no more likely to compromise their agency than is the natural history of those capacities that constitute agency. From this perspective spontaneity and its accompanying identification with non-rational features of the self appears more like a fulfilment of human nature than a departure from it.

§ 5. Conclusion

I have argued that Korsgaard's integrity theory cannot accommodate uncabined spontaneity nor can it accommodate some of the paradigmatically valuable activities associated with spontaneity—those activities in which we see ourselves as outstripping our rational natures. At the same time, selves capable of experiencing spontaneity cannot accommodate some of the values that are 'baked in' to the integrated Korsgaardian self, such as the guarantee of non-alienation. Reflectively integrated selves can come close to experiencing some of the values associated with spontaneity in the creativity of interpreting principles and practical identities, and selves capable of experiencing spontaneity can come close to achieving the values associated with integrity by, for instance, lucking into a weak sort of non-alienation. But neither sort of self can fully achieve the values associated with the other sort. This suggests that the incompatibility between Korsgaard's integrity theories of agency and my account of spontaneity is a deep discordance, not a surface level conflict: the values achieved by experiencing spontaneity can conflict with those achieved by self-regulation through one's rational, deliberative perspective.

Reflective integrity theorists such as Korsgaard argue that the view that successful human lives necessarily involve agential self-constitution is entailed by a transcendental argument: given that we, in fact, ask questions about the best way to live our lives, we should adopt integrity theory's view of the self on pain of inconsistency. I have answered Korsgaard's transcendental argument by offering a series of counterdemonstrations: many people do, in fact, make themselves into less integrated agents whilst remaining intelligible interlocutors and subjects of experience. I have not argued definitively that this picture of the spontaneous self is, ultimately, the best picture to adopt, although I incline toward it. Because the two pictures of the self exclude each other, we are left with a potentially insoluble philosophical problem. We might attempt a pluralistic compromise—for instance,

regarding each theory of the self as a describing an important but different aspect of the self—but even such an attempt at compromise would itself require rejecting Korsgaard’s transcendental argument and the demand for rational unity that motivates reflective integrity theories of the self.¹³

References

- Arruda, Caroline T. 2016. ‘Constitutivism and the Self-Reflection Requirement.’ *Philosophia* 44: 1165-83.
- Bagley, Benjamin. 2017. ‘Properly Proleptic Blame.’ *Ethics* 127: 852-82.
- Bishop, Elizabeth. 1979. ‘Sonnet.’ *New Yorker*. 29 October.
- Bratman, Michael E. 2007. *Structures of Agency: Essays*. Oxford: Oxford University Press.
- Bratman, Michael E. 2012. ‘Time, Rationality, and Self-Governance.’ *Philosophical Issues* 22: 73-88.
- Crowell, Steven. 2007. ‘Sorge or *Selbstbewußtsein*? Heidegger and Korsgaard on the Sources of Normativity.’ *European Journal of Philosophy* 15(3): 315-33.
- Dreyfus, Hubert L. 1993. ‘Heidegger’s Critique of the Husserl/Searle Account of Intentionality.’ *Social Research* 60(1): 17-38.
- Dworkin, Ronald. 2011. *Justice for Hedgehogs*. Cambridge, MA: Harvard University Press.
- Elliott, Carl. 2003. *Better Than Well: American Medicine Meets the American Dream*. New York: Norton.
- Jaeggi, Rahel. 2014. *Alienation*. Edited by Frederick Neuhouser. Translated by Frederick Neuhouser and Alan E. Smith. New York: Columbia University Press.
- Gingerich, Jonathan. n.d. ‘Spontaneous Freedom.’ Unpublished manuscript.
<https://jonathandgingerich.files.wordpress.com/2021/01/spontaneous-freedom-2021-01-01.pdf>.
- Grob, Charles S. 2007. ‘The Use of Psilocybin in Patients with Advanced Cancer and Existential Anxiety.’ In *Psychedelic Medicine: New Evidence for Hallucinogenic Substances as Treatments*, vol. 1, edited by Michael J. Winkelman and Thomas B. Roberts, 205-16. Westport, CT: Praeger.
- Korsgaard, Christine M. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Korsgaard, Christine M. 1999. ‘Self-Constitution in the Ethics of Plato and Kant.’ *Journal of Ethics* 3: 1-29.
- Korsgaard, Christine M. 2008. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.
- Korsgaard, Christine M. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- Millgram, Elijah. 1997. *Practical Induction*. Cambridge, MA: Harvard University Press.

¹³ Many thanks are due to Benjamin Bagley, Daniela Dover, Jane Friedman, Barbara Herman, Pamela Hieronymi, A.J. Julius, Jordan MacKenzie, C. Thi Nguyen, Calvin Normore, Mike Ridge, Nick Schwieterman, Seana Shiffrin, an anonymous referee, and audiences at the UCLA Ethics Workshop and APA Eastern for helpful comments on earlier versions of this essay.

- Nagatomo, Shigenori. 2017. 'Zen Buddhist Philosophy.' In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring.
<https://plato.stanford.edu/archives/spr2017/entries/japanese-zen>.
- Nguyen, C. Thi. 2019. 'Games and the Art of Agency.' *Philosophical Review* 128(4): 423-62.
- Pauer-Studer, Herlinde. 2018. 'Korsgaard's Constitutivism and the Possibility of Bad Action.' *Ethical Theory and Moral Practice* 21: 37-56.
- Pollan, Michael. 2018. *How to Change Your Mind: What the New Science of Psychedelics Teaches Us about Consciousness, Dying, Addiction, Depression, and Transcendence*. New York: Penguin.
- Ridge, Michael. 2019. 'Play and Games: An Opinionated Introduction.' *Philosophy Compass* 14(4): e12573.
- Strawson, Galen. 2009. *Selves: An Essay in Revisionary Metaphysics*. Oxford: Oxford University Press.
- Velleman, J. David. 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press.
- Velleman, J. David. 2006. *Self to Self: Selected Essays*. Cambridge: Cambridge University Press.
- Velleman, J. David. 2009. *How We Get Along*. Cambridge: Cambridge University Press