INTEGRITY AND SPONTANEITY

Jonathan Gingerich

October 22, 2018

§ 0. Abstract

Phenomenologically, experiencing myself as spontaneously free involves regarding the future of my life as settled by neither internal, reflectively endorsed features of me nor the decisions of other agents. Theories of the self that require that I constitute myself through rational deliberation and planning in order to exist as a self preclude experiences of spontaneous freedom. I argue that while such "integrity" theories of the self, in their most stringent forms, have a variety of attractive features, their attractiveness depends on constructing a sort of self that cannot experience and value spontaneous freedom. I further contend that transcendental arguments that purport to show that simply by philosophizing about the self we commit ourselves to a stringent integrity theory are uncompelling because of our ability to talk with and about spontaneous actors without difficulty. The value of spontaneous freedom ultimately gives us reason to adopt a less rationalized, more embodied theory of the self than that implicitly advanced by stringent versions of integrity theory.

§ 1. Introduction

Major strains in moral philosophy extol the importance of integrity and coherence for agency, claiming that for a person to function effectively as a rational agent, or even to qualify as a rational agent, the person must exhibit integrity and must be coherent to themselves and others. Some stringent versions of *integrity theory* maintain that successful selves are selves whose heteronymous, foreign elements have either been weeded out or subjugated. A prominent version of integrity theory is advanced by Christine Korsgaard, who maintains that successful agency requires s*elf-constitution* through reflective endorsement of—and action in accordance with—*practical identities* that provide laws to guide agents' choices and actions. For Korsgaard, I appropriately relate to myself by self-constituting through my rational capacities.

The "self-constitution" theory promulgated by Korsgaard produces an account of agency according to which moral agents who successfully self-constitute are guaranteed to endorse their actions and beliefs, can make promises and agreements that they keep out of

more than a *modus vivendi*, and commit themselves to a Kantian morality. However, in spite of these attractions, the demands of integrity theory of the stringent variety advocated by Korsgaard—that agents make themselves coherent and expel foreign elements that undermine self-coherence—preclude agents from experiencing *spontaneous freedom*—the freedom of a free spirit, or someone who is "free as a bird." To experience spontaneous freedom is to feel that one's actions are settled in advance neither by one's own conscious, deliberative plans and decisions or those of other agents.

In this essay, I describe integrity theory and explain why its ethical view of the self is, at first glance, attractive. I then show that integrity theory is incompatible with the experience of spontaneous freedom. I argue that Korsgaard's theory accommodates a great deal of creativity but ultimately precludes some of the central features of spontaneous freedom—features that are part of why spontaneous freedom is worth wanting. I argue that one cannot experience spontaneous freedom when one's actions are settled in advance by one's consciously and reflectively endorsed practical identities. The incompatibility between spontaneous freedom and stringent integrity theories is not just a surface-level tension but a deep and intractable conflict about the nature of the self that turns on the question: can I see myself in my subconscious, non-deliberative features or not?

§ 2. The Integrated Self

In *Self-Constitution: Agency, Identity, and Integrity*, Korsgaard sets out to show that, in order 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 an 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" (Korsgaard 2009, 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" (Korsgaard 2009, 20). These practical identities all provide "absolute inviolable laws" to guide an agent's choices (Korsgaard 2009, 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" (Korsgaard 2009, 21-22).[1] A good action "is one that both achieves and springs from the integrity of the person who performs it" (Korsgaard 2009, 25).

Achieving integrity further requires agents to resolve conflicts among their practical identities when the principles or laws provided by those identities come into conflict, as a failure to do so would inevitably result in failing 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 (Korsgaard 2009, 72-76). "Universal willing" can be "provisionally universal" for

[1] 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).

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.[2]

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" (Korsgaard 2009, 1). Insofar as you aim to have reasons for how you act and live, you are committed to seeking universally applicable reasons for how you can act and live as a single and unified individual. 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 (Korsgaard 2009, 15-16), so being a good agent is the only thing that is choiceworthy for a person (Korsgaard 2009, 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.

---

[2] This example is meant only to illuminate the distinction between universal and particularistic willing. Korsgaard might well regard it as 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.

A successful human agent is neither a "Good Dog," who "always does what he ought to do spontaneously and with tail-wagging cheerfulness and enthusiasm," nor a "Reformed Miserable Sinner," who constantly experiences unruly and discordant desires and impulses that he "must constantly repress … in order to conform to the demands of duty" (Korsgaard 2009, 3). However, the Reformed Miserable Sinner picture is closer to the mark, for Korsgaard, than is the Good Dog. Like the Reformed Miserable Sinner, Korsgaard's ideal agent represses her unruly desires; the difference is that Korsgaard's successful agent represses threats to her integrity "in order to be one, to be unified, to be whole" rather than "to be good" (Korsgaard 2009, 26). 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.

A recovering alcoholic asked out to a bar by a few friends might exhibit integrity by deciding not to go, not because it would be "bad" to have a drink but because doing so would violate the recovering alcoholic's practical identity as a teetotaler. For Korsgaard, an ongoing, unresolved conflict between practical identities is incompatible with success at self-constitution. The path to integrity, for Korsgaard, is not to allow the uneasy coexistence of two identities providing conflicting guidance. An agent must fit their identities together so that they do not *potentially* conflict, otherwise that agent's capacity for effective (and unified) action is at the mercy of accident. Someone whose practical identities provide conflicting principles or laws can only contingently live up to their roles until such time as they jettison or revise one or more of their identities to resolve the conflict.

The recovering alcoholic might also achieve integrity with different decisions. Perhaps the best way to follow the principles provided by the practical identities of "teetotaler" and "friend" is to go to the bar and just drink seltzer water, or perhaps the temptations of alcohol

have receded sufficiently far that the recovering alcoholic has reason to revisit her identity as a teetotaler—maybe one glass of beer is fine. But as long as the agent reflectively endorses the identity of "teetotaler," the agent is committed to living up to it. For Korsgaard, "so long as you remain committed to a role, and yet fail to meet the obligations it generates, you fail yourself as a human being, as well as failing in that role" (Korsgaard 1996, 121).

To live up to the practical ideal of integrity, according to Korsgaard, such a change in practical identity must be *principled* rather than capricious. An agent might succeed at self-constitution by adopting a practical identity that calls for artistic spontaneity or for fun, but for a good person on Korsgaard's picture, such self-constitution cannot involve shifting one's commitments without reason. "[Y]ou have to will universally, because the reason you act on now, the law you make for yourself now, must be one you can will to act on again later, come what may, unless you come to see that there's a good reason to change it" (Korsgaard 2009, 202-03). This contrasts with the "particularistic will" of a person who "expects to change his mind without a reason"; such a person "lack[s] self-respect" and is not governed by the law of their own will (Korsgaard 2009, 203).

Korsgaard's integrity theory of the self is potentially attractive in that selves that satisfy the demand of self-constitution have several valuable features.

First, 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). The agent who successfully self-constitutes comes to see all of their life as their own, because they reflectively endorse those practical identities that guide their actions and jettison those that cannot fit with their other identities.

Second, success at Korsgaardian self-constitution enables agents to keep promises and agreements out of rational necessity, rather than "contingently." To make promises successfully, Korsgaard thinks, one must be able to make *necessary* that one will live up to the promise in the future. This can be accomplished by those people, and only those people, who make themselves into single agents through self-constitution. "Mere heaps" are unable to make real promises because the part of them that makes the promise might not be the part of them that takes control when it is time to live up to the promise. If I promise my neighbor that I will help paint their house but do so without successfully self-constituting—say, I have an unresolved conflict between my identity as a surfer and my identities as a promise keeper and as a neighbor, whether I keep my "promise" will just depend on which part of my identity is ascendent when the first sunny day comes (Korsgaard 2009, 22). Heteronomous heaps might keep "promises" out of a *modus vivendi*: a contingent peace among their conflicting, unintegrated identities. But such "promise-keeping" does not succeed at "making the contingent necessary." Korsgaardian agents, in contrast, have the ability to keep promises "in the noumenal world" (Korsgaard 2009, 191).[3]

§ 3. Integrity and Spontaneity

Spontaneous freedom is a sort of freedom that involves activity that outstrips a self's conscious identities in the sense that the grounds of the spontaneous activity are not their conscious, first-personal reflection and endorsement. In this section, I argue that selves that satisfy Korsgaard's demands for successful agency cannot experience spontaneous freedom,

[3] On Korsgaard's view, successfully self-constituting agents not only have the capacity to make and keep agreements and promises in the noumenal world but also have a substantive commitment to the norms of interpersonal morality (Korsgaard 2009, 181).

because the integrity and self-understanding achieved by such agents requires that their full-fledged actions be ascribable to practical identities that they reflectively endorse.

Korsgaard's view requires that practical identities provide us with laws that guide our actions, and for Korsgaard "[a] law lays down what is to be done" (Korsgaard 2009, 16). Acting as a successful agent in accord with a practical identity involves an experience of "necessitation," which involves "work and effort" (Korsgaard 2009, 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.

Experiences of spontaneous freedom 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 might agree that stringent requirements for agential integrity rule out the possibility of experiencing spontaneous freedom, but might argue that integrity still provides plenty of room for creativity and spontaneity and that whatever is *valuable* about spontaneous freedom is provided for just as well by the forms of creativity and spontaneity that they allow. 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). I might adopt a

practical identity that says that when I am creating art, I should create whatever I am moved to create, not for any instrumental reason but just for the pleasure of spontaneous artistic creation. Korsgaard also provides space for spontaneity in interpretation, argument, and creativity in determining what any particular 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" (Korsgaard 2009, 21). So, 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.

While the forms of creativity that are accommodated by integrity theory are important and valuable, I will contend in this following section that some of the most paradigmatically valuable experiences of spontaneous freedom cannot be had by selves that satisfy the demands of Korsgaard's integrity theory.

§ 4. Art, Fun, and Games

In this section, I argue that the activities of making creative art, playing games, and having fun depend on taking your actions not to be determined by practical identities that provide action-guiding laws. Consider the three following spontaneously free "characters," each of which exemplifies a different side of the ideal of spontaneous freedom:

First, consider a creative artist who regards their artistic creativity as involving creation that is not guided by any rule or determinate principle. The artist might agree with Kant that "one cannot learn to write inspired poetry, however exhaustive all the rules for the art of poetry and however excellent the models for it may be" because no successful creative artist "can indicate how his ideas, which are fantastic and yet at the same time rich in thought, arise and come together in his head, because he himself does not know it and thus cannot teach it to

anyone else either" (Kant 2000, 5:309). For this artist, artistic creation involves activity that they might experience as the world "flowing" through them, rather than activity guided by principles or laws to which they subscribe. This artist might create in accordance with an imperative that says, "Create!" but the imperative must be so indeterminate that it does not meaningfully specify what the artist should do. Such an artist is committed to thinking that there is some part of the self that they cannot know until it comes to the surface, so that even they might be surprised by who they are and what they will do.

Second, consider a game player. According to Bernard Suits, a game is the "voluntary attempt to overcome unnecessary obstacles" (Suits 1990, 41). A Suitsian game player regards the activity of playing a game (although not the decisions that one makes within a game while playing it) as not necessitated in advance and as non-instrumentally worthwhile. A common theme of philosophical discussion of games, both those associated with Sutis's attempt to define games and those that adopt more pluralistic views about what counts as a game, is the point that playing games is often or always associated with a feeling of non-obligatoriness. As Roger Caillois argues, if playing were obligatory "it would at once lose its attractive and joyous quality as diversion" (Caillois 2001, 9). The activities that a game player performs *within* a game's "magic circle"—the space that players inhabit when they play a game—are constrained by the rules of the game and the player's aim of winning the game (Huizinga 1950, 10). When I play chess, my ability to move my rook is constrained by the rules (it cannot move diagonally) and my aim of winning (I cannot move my rook to a square where it can be costlessly captured by my opponent if I am really trying to win). I might play a game of chess in order to spend time with a friend and exercise my spatial reasoning skills. However the activity of *playing* a game is itself necessarily experienced by players as non-obligatory. An activity that is

recognizable as "playing a game" is an activity that the game player does not *have* to do, and that the player regards as such. In this respect, an experience of spontaneous freedom is provoked by the playing *of* any activity recognizable as a game. Part of what it is to play a game is to regard the activity of playing the game as itself optional and not required, even by any self-legislated rule or law.

Play understood more broadly than game-play also provides an experience of non-instrumental, voluntary, intrinsically valuable activity (Tasioulas 2004, 244). Consider, third, someone "having fun," where "having fun" is an activity that is, necessarily, undertaken just for its own sake, rather than for any further reason. Johan Huizinga regards "the fun element" as precisely what "characterizes the element of play," suggesting a strong continuity between the practical identity of the game player and the person who has fun (Huizinga 1955, 8). I stipulatively take "fun" to be non-rule constituted play activity (i.e., play that is not a game). "Fun" is different than intrinsically valuable activity and often involves the activation of interest or attention in new directions that are not specified antecedently to the activity of "fun." Fun involves "being in the moment," finding the moment in which one sees oneself satisfying or pleasing, and much of the time involves feeling excited anticipation arising from uncertainty about what, exactly, one will do next. Fun involves the free movement of interest or attention and depends heavily on the attitudes of the participants. Although it is difficult to describe specific examples of fun out of context—fun is often something for which "you had to be there"—many activities like making up a game, having a conversation, flirting, or making dinner can be fun or not depending on the attitudes of the participants. In contrast to the activity of game playing, fun is not rule-governed. Someone having fun often regards the next

place they will turn their attention, the next thing they will do, as not specified in advance by any rule. This non-specification gives fun its characteristic feeling of excitement.[4]

If we think of "character" as a "formal device that collects every example of a kind of person," (Kunin 2009, 291), the experiences associated with artistic creativity, playing games, and having fun might be understood as instances of character types, or recognizable "kinds" of person that an individual might be, for some temporal period, if not for the entirety of their life.[5] We have all encountered persons for whom spontaneity or fun is central to who they are. A person might inhabit one of these character types for only a portion of their life, leaving it behind when other practical identities become more salient to them, although leaving behind the character type might not mean abandoning the associated practical identities. The formerly "wild artist" who now spends all of their time working as a lawyer and caring for their kids might still hold on to the practical identity of "artist," and might experience their inability to satisfy that identity in addition to the identities of "professional" and "parent" as a source of frustration.

Korsgaard might wish to regard the "spontaneous artist," "game player" and "fun lover" 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 ends up

---

[4] I do not mean to claim that every experience of fun *must* involve spontaneity. One might experience fun in a different sense through engagement and total immersion in an activity without distraction, as when one is completely absorbed in reading a book. The sort of fun that involves the non-rule-governed fixation of attention is a particularly piquant form of fun because this feature seems to account for the boisterousness associated with fun in ordinary language. Thanks to <anonymized> for raising this point.

[5] A life made up *entirely* of playing games—at least in a world characterized by the injustices that characterize our own world rather than the sort of utopia in which Suits imagines playing games would be the only thing for us to do (Suits 1990, 168)—would feel empty or shallow. This is both because the value of experiencing spontaneous freedom should not be thought of as the only value that gives meaning to a life and because the sort of spontaneous freedom provided by playing games is more restricted than that provided by artistic creativity and non-rule-bound fun.

suggesting that there is something defective in all three of these spontaneously free character types. For Korsgaard, spontaneity must be suitably cabined so as not to give rise to intra-agential conflict. For Korsgaard, "[I]f you expect to change your mind without a reason, then you are not willing your maxim as a universal law, not even a provisionally universal law.... And if you aren't willing your maxim as a universal law, then you lack self-respect" (Korsgaard 2009, 203). Similarly, "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). But this is precisely what an artist who experiences their creativity as "nature flowing through them," does: they understand their agential role as one of observing and recording motivational forces that arise in them without their active participation. Likewise, the game player and fun lover identify with motivational currents that are part of "who they are" but that are not accessible to conscious reflection. The fun lover has fun by allowing their attention to move from object to object with no explicit, reflective guidance.

Korsgaard might *partially* accommodate the spontaneity of these characters. Perhaps the spontaneous action of the artist can be described at a higher level of abstraction. What is "to be done" is create spontaneously. The particulars of the art that gets created are, perhaps, not part of what needs to be specified by the principles given by practical identities. An artist might create one sort of painting today simply because it "feels like the right combination of images" and create a dramatically different painting tomorrow for the same reason without giving rise to particularistic willing because the artist's willings are temporally limited to the day that they occur, so no conflict between the two willings arises. Such an artist does not

expect to change their mind without a good reason. Rather, they create spontaneously today and then again tomorrow.

However, such an artist would lack a Korsgaardian practical identity, because such interpretations are too vague to qualify as action-guiding laws or principles; they have been constructed precisely to *avoid* specifying what is to be done by the agent. Korsgaard is clear that successful agents must have *some* practical identity that prescribes what to do. "What is not contingent is that you must be governed by some conception of your practical identity," Korsgaard explains, "[f]or unless you are committed to some conception of your practical identity, you will lose your grip on yourself as having any reason to do one thing rather than another—and with it, your grip on yourself as having any reason to live and act at all" (Korsgaard 1996, 120-21).

Integrity theory cannot allow the artist's spontaneity to interfere with their other practical identities. If the artist 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, the person's artistic spontaneity cannot be realized 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 the identity as a creative artist has only as much free play as is permitted by all of the artist's other practical identities. Adopting a practical identity that authorizes spontaneity is compatible with successful self-constitution if the spontaneity is suitably cabined. On Korsgaard's view, a well-constituted agent's deliberative, consciously endorsed practical identities authorize 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). But much of the value of experiencing spontaneous freedom is associated with temporarily turning off or escaping from these rational control systems for one's 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. In Korsgaard's view, if the agent's only or most important practical identity involved being spontaneous, they would be "at the mercy of accident" and "almost completely *incapable of effective action*" (Korsgaard 2009, 169) since their actions would not be necessitated by a practical identity. 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.

The artist might object that their spontaneity is different from "living at random," both because they endorses their spontaneity and because their spontaneity does not involve acting completely at random. Rather, they spontaneously choose among different practical identities that they care about. It is one's unique material and social circumstances and one's psychological states other than plans and decisions that produce spontaneous activity, not the role of a die.[6] But, Korsgaard would point out, this still makes the artist's success *contingent*. It is only *if* their spontaneity does not cause them to careen from activity to activity—which it might—that they can accomplish *anything*.

---

[6] We might worry that making all of our important decisions with a throw of the die would make them spontaneous but not in the right way. Whether this is so depends on how we might decide using a die. If I decided to set up all of my decisions in sets of six options and to commit myself to whichever course of action the roll of a die indicated, I would not be spontaneously free in my action because my rule-like commitment to doing what the die told me to do would motivate my action. On the other hand, it seems that I could be right to regard my own spontaneity as, in some way, *analogous* to a die roll. Part of the pleasure of experiencing spontaneous freedom is seeing myself as made up all of the natural elements that contingently happen to make me up. Thanks to <anonymized> for pressing this point.

Even if the artist could avoid agential failure by adopting a practical identity involving spontaneity, achieving Korsgaardian integrity would depend on subordinating their other practical identities at issue to this identity. If they did not subordinate these identities to their identity as someone who is spontaneous, they would still be subject to criticism on the grounds that their identities conflict with one another: their spontaneous identity says they should keep working on the sculpture while their identity as a punctual friend says to put down the chisel. Only by having some procedure that they could will universally, such as a procedure that prioritizes the spontaneous identity above the other potentially conflicting identities while at the same time suitably restricting its scope so that they avoid living "at random," could the artist avoid the internal conflict characteristic of someone who fails to self-integrate and so is a "mere heap."

Korsgaard might claim that possessing conflicting practical identities remains *possible* for humans on her view, even though holding on to conflicting practical identities represents a failure of agency, since integrity is a practical *ideal* for Korsgaard. But Korsgaard's theory of successful agency purports to provide an account of what it is to be a *good human*. This is not to suggest, for instance, that anyone should be legally compelled or coerced to integrate their practical identities. But on Korsgaard's view, being a *good* human is incompatible with experiencing spontaneous freedom insofar as such experiences are incompatible with successful agency, which all humans are committed to pursuing as far as possible.

I have argued that the incompatibility between integrity theory and the possibility of experiencing and valuing spontaneous freedom arises from an incompatibility between the ethically successful self envisioned by integrity theory and the values realized by experiences of spontaneous freedom. This incompatibility runs the other way, too. As I have already

suggested, selves that can incorporate sources of action that are not suitably authorized by their "constitutions" do not have the power to keep promises "in the noumenal world," for their capacity to keep promises depends on which of their practical identities has the upper hand when their promise comes due.

However, pragmatically speaking, heteronymous, spontaneously free selves might keep their word almost as well as integrated selves. In considering the neighbor's request for help painting a fence on the first sunny day, a spontaneous actor might predict what motivational state they will be in when the first sunny day comes. If they predict that they will feel the bonds of neighborly obligation tugging more strongly than the lure of the beach on that day, they might go ahead and "promise" to help paint, knowing that there is some chance that their competing identities will win out on the first sunny day but thinking that the chance is low. When the first sunny day comes, they will, more likely than not, end up helping their neighbor paint. The neighbor making the quasi-promise might be more likely to actually end up showing up to help paint than the promisor in the noumenal realm, for even the "true" promisor might not show up to help paint if sufficiently morally weightily countervailing considerations come into play or if their identities prove to be less integrated than they thought they were. Yet, on Korsgaard's view, the quasi-promisor has not really succeeded at promising, because in making their "promise," they did not rationally necessitate that they would help their neighbor paint. But the quasi-promise that they made is, as far as the neighbor is concerned, almost as good as a promise "in the noumenal world."

There are other 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.[7] 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. This would count as the sort of "unprincipled resolution" that I have identified as characterizing spontaneous activity, provided that such a theory allow "principled-ness" to encompass even activity arising from one's identification with momentary, non-rational impulses and instincts. Because experiences of spontaneous freedom can arise from epistemic uncertainty, as well as from metaphysical non-determination, the sort of principled integrity that allows or requires a strong form of discretion in acting out one's principles allows ethically successful agents to experience spontaneous freedom, unlike the view advanced by Korsgaard.[8]

§ 5. Korsgaard's Transcendental Argument

A further argument remains in support of Korsgaard's stringent version of integrity theory: a transcendental argument that by philosophizing 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 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,

---

[7] 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).

[8] For an account of weak and strong forms of discretion, see Dworkin 1977, 31-39.

we must adopt a different understanding of the nature of the self than that advanced by Korsgaard. A self is capable of experiencing spontaneous freedom 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 advances a "transcendental argument" in favor of her self-constitution theory, claiming that by seeking justifications for your actions, you commit yourself to being a single "you," which you can only accomplish by having some 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. Korsgaard's transcendental 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 do the hard work of falling apart and pulling yourself back together, which requires that you decide how your identities cohere rather than just flitting from one to another for no reason at all *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 are animals who need practical conceptions of their own identities in order to find their actions worth undertaking, humans

are committed to taking their practical identities to be "normative," rationally necessitating 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 transcendental argument is to point out that experiences of spontaneous freedom require temporarily *turning off* the intellectual drive that seeks reasons to act and to live. Experiences of spontaneous freedom 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 deliberative control of their actions that characterizes spontaneous freedom make use of the connection between conscious self-awareness and the physical body. Consider two examples.

First, a long 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 is arrested the activity of an individual practitioner's ego-consciousness" (Nagatomo 2017, § 6.3). This experience is brought about, in substantial part, though 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" (Nagatomo 2017, § 8.2). Action then "carries a sense of *spontaneity*" that surges "from the creative source in the bottomless ground" (Nagatomo 2017, § 8.2).

Second, psychiatrists have developed an interest in studying the use of hallucinogens to treat depression and existential anxiety in patients with late-stage cancer (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 experience a temporary failure of the impulse for individuation. When this experience takes place in the context of supportive psychotherapy or nurturing relationships, subjects can connect this experience of the failure of individuation to their sense of self after the hallucinogenic session and can come to see themselves as less fully identified with their own rational impulses and so 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 art, or a game, or having fun can have a similar effect. Temporarily severing the connection been self-awareness and theoretical intelligence often allows experiences of spontaneous freedom 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 entirely 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-transcendental 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 there existing some single agent to whom such actions can be attributed. How can such an agent exist, other than through self-constitution? Here, Korsgaard might turn to her account of defective action. Consider Korsgaard's story about 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).[9]

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

[9] 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).

into drinking his beer. For Korsgaard, Jeremy can only succeed as an agent by acting *non-contingently* (Korsgaard 2009, 169).

Jeremy might reply to Korsgaard's 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 (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."[10] Korsgaard asks rhetorically, "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). A city constituted by interest groups that have simply made pragmatic alliances with one another is

[10] For an account of a self that encompasses more than its conscious, deliberative nature, see Hubert Dreyfus's reading of Heidegger on intentionality ("[P]henomenological examination shows that in a wide variety of situations human beings relate to the world in an organized purposive manner without the constant accompaniment of a representational state which specifies what the action is aimed at accomplishing" (Dreyfus 1993, 23-28).)

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 self-hood 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, the 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 in order to experience spontaneous freedom.

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. Korsgaard would, then, be better off giving up her transcendental argument than her account of defective agency. She could then maintain that Jeremy is a defective agent without claiming that he is thereby a defective *human*.

Why is being a rational agent the only way of successfully being human? Why cannot one shift from identity to identity in an unprincipled manner? Korsgaard's integrity theory maintains that incomplete self-constitution is *defective* rather than effective but partial and regards reliance on instinct or mere inclination as human failures *tout court*. 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).

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 "reasons's own principle," is contingent and unstable. We might adopt a view of the mind and it's sense of coherence and order like that of Freud who, in Francey Russell's interpretation "provides a historical-developmental account of the conditions for psychic order, which is to say, paradoxically, that for Freud the conditions of possible experience are themselves conditioned" (Russell 2012, 379). 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. Likewise, if one thinks that artistic creativity and playfulness are just as essential to humanity as is rationality, spontaneity and its accompanying identification with non-rational features of the self appears more like a fulfillment of human nature than a departure from it.

§ 6. Conclusion

I have argued that integrity theory cannot accommodate spontaneous freedom nor can it accommodate some of the paradigmatically valuable activities associated with experiences of spontaneous freedom: making and appreciating creative art, playing games, and having fun. At the same time, selves capable of experiencing spontaneity cannot accommodate some of the values that are "baked in" to the integrated Korsgaardian self: the capacity to rationally necessitate one's actions through promises and agreements and the ruling out of self-alienation. Integrated selves can come close to experiencing some of the values associated with spontaneous freedom in the creativity of interpreting principles and practical identities, and selves capable of experiencing spontaneous freedom can come close to achieving the values associated with integrity by, for instance, contingently keeping their word. But neither sort of

self can fully achieve the values associated with the other sort. This suggests that the incompatibility between stringent integrity theories of agency and my account of spontaneous freedom is a deep discordance, not a surface level conflict: the values achieved by experiencing spontaneous freedom can conflict with those achieved by self-regulation through one's rational, deliberative perspective.

Integrity theorists such as Korsgaard further 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 counter-demonstrations: many people do, in fact, make themselves into less integrated agents while 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. Insofar as the Korsgaardian self has the power to keep promises "out of necessity" while the spontaneous self can have fully-formed experiences of artistic creativity, play, and fun, the two pictures of the self serve incommensurable values. Because the two pictures of the self rule each other out, we are left with a potentially indissoluble 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 require rejecting Korsgaard's transcendental argument and the demand for rational unity at the heart of stringent integrity theories. [11]

[11] Many thanks are due to Daniela Dover, Thi Nguyen, Jane Friedman, Nick Schwieterman, Seana Shiffrin, Calvin Normore, Pamela Hieronymi, A.J. Julius, and Barbara Herman for helpful comments on earlier versions of this essay.

REFERENCES

Bratman, Michael E. 2012. "Time, Rationality, and Self-Governance." *Philosophical Issues* 22: 73-88.

Caillois, Roger. 2001. *Man, Play and Games*. Translated by Meyer Barash. Urbana: University of Illinois Press.

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. 1977. *Taking Rights Seriously*. Cambridge, MA: Harvard University Press.

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: W.W. Norton and Company.

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.

Huizinga, Johan. 1955. *Homo Ludens: A Study of the Play Element in Culture*. Boston: Beacon Press.

Jaeggi, Rahel. 2014. *Alienation*. Edited by Frederick Neuhouser. Translated by Frederick Neuhouser and Alan E. Smith. New York: Columbia University Press.

Kant, Immanuel. 2000. *Critique of the Power of Judgment*. Edited by Paul Guyer. Translated by Paul Guyer and Eric Matthews. Cambridge: Cambridge University Press.

Korsgaard, Christine M. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Korsgaard, Christine M. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Kunin, Aaron. 2009. "Characters Lounge." *Modern Language Quarterly* 70(3): 291-317.

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.

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.

Russell, Francey. 2012. "Unity and Synthesis in the Ego Ideal: Reading Freud's Concept through Kant's Philosophy." *American Imago* 69(3): 351-81.

Suits, Bernard. 1990. *The Grasshopper: Games, Life, and Utopia*. Boston: David R. Godine, Publisher.

Tasioulas, John. 2006. "Games and the Good." *Proceedings of the Aristotelian Society*. Supplementary Volume 80: 237-64.