

## **The Virtue of Integrity: A New Framework and Integrative Review** **Dr. Matthew Kuan Johnson and Dr. Rachel Siow Robertson**

This is an unpublished conference paper for the 10<sup>th</sup> Annual Jubilee Centre for Character and Virtues conference at Oriel College, Oxford University, Thursday 8<sup>th</sup> – Saturday 10<sup>th</sup> September 2022. These papers are works in progress and should not be cited without author's prior permission.

**Abstract:** Despite the vast literature on the virtue of integrity and its importance to practitioners, there is little consensus regarding what integrity actually is. Our paper proposes a novel definition of integrity and demonstrates the advantages of this new approach for organizing the conceptual space around the virtue, explaining associated psychological phenomena, and for practitioners seeking to promote integrity. The second part of our paper outlines the structural conditions necessary for integrity, to provide an account of how integrity is developed. We conclude by drawing from Medina’s (2013) work on “meta-lucidity” to explore an interesting (apparent) paradox raised by our account.

## 1. Introduction

Integrity is one of the most important virtues, with some arguing that it is a necessary condition for “constituting the self” (Frankfurt, 1987) or even moral agency (Korsgaard, 2009). Practitioners also make frequent use of the concept of integrity, including in approaches to management and organizational studies (e.g. administering pre-employment integrity tests, policing of integrity in sports management, and attempts to cultivate integrity within military ethics; see Palanski & Yammarino 2007), education (e.g. testing and promoting the academic integrity of students), and therapy (e.g. aiming at the integration and acceptance of the self in psychotherapy). For these reasons, there is a vast literature on the concept in philosophy and psychology. However, approaches within both fields tend to cluster around very narrow definitions of integrity, often focusing on either the epistemic or moral form of the virtue, or focusing on components that make up the virtue, such as honesty, self-efficacy, and avoiding moral hypocrisy, rather than on the overall virtue itself. What is needed, then, is an overarching framework that can situate all of these narrowly defined components of integrity into a broader understanding of the virtue itself, that will be of use to philosophers, psychologists, and practitioners.

Some of the most pressing questions to answer with respect to integrity are illustrated by the following test cases.

### 1.1. Morality and Integrity (problem cases)

What are the connections between morality and integrity? One major area of debate has to do with whether integrity is compatible with being immoral or committing immoral acts. Problem cases that are usually discussed include:

- *Being fully committed to (and consistent in) doing harm.* Consider how Hitler was apparently well integrated in that his beliefs, desires, and behaviours were consistent with one another, as they were integrated around harming certain groups of people. Consider also Milton’s Satan, who famously declared ‘Evil be thou my good’ (Milton 1667/2000,

see IV: 110): such a figure is also apparently well integrated around their commitment to pursuing evil (see Johnson 2020a for a discussion).

- ***Doing harm in the course of being committed to some project.*** Consider Bernard Williams' (1981: 23) famous case of Gauguin leaving his family to pursue his art, spies who have to lie, the person who commits adultery in the name of love, etc.
- ***Making a choice when faced with a moral dilemma.*** One class of moral dilemmas that makes this case particularly strongly are what Lisa Tessman refers to as 'tragic dilemmas': these are cases in which while there may be an all-things-considered *best* thing to do, doing the *right* thing is impossible (Tessman 2015, 2017). Tessman relates how being placed in a moral dilemma precludes the possibility of a sense of integrity, as one experiences being pulled in contrary directions prior to the moment of decision, and after the moment of decision experiences a kind of 'moral remainder' (which threatens their sense of integrity) from the knowledge that they chose an option they they knows was not the *right* (in an unqualified sense) thing to do.
  - ***Cases from Bernard Williams*** (1973: 98-100).
    - *The "Jim and the Indians" case:* Jim and 20 others have been kidnapped by a deranged individual who will kill all 20 unless Jim shoots one of them.
    - *The George the Chemist case:* George is having difficulty finding a job, and needs to be able to provide for his family. His only job possibility is working on chemical warfare, which he is morally opposed to; however, if he takes the job, he can mitigate some of the bad effects of the job which the other individual who would take the job would not do.
  - ***Acts committed in situations of systemic injustice,*** such as some forms of political protest/resistance, where certain commitments are broken (e.g. commitments to the laws or norms of the land, such as not disrupting traffic with one's protest or following racist bus seating protocols), in order to honor other commitments (e.g. to work for the liberation of one's community).

There are also cases in which there may be integrity despite inconsistency or lack of commitment. This includes "behavioural inconsistency" (inconsistency between commitments and behaviour, as documented by Monin & Merritt 2012):

- ***Preaching without practicing:*** e.g. a doctor who has a duty to recommend that his patients do not smoke, even if he, himself, is a heavy chain smoker.
- ***Not practicing and then preaching:*** e.g. a case of redemption where one commits to and extols the virtues of clean living, after having spent the past decade of their life addicted to drugs and hit their rock bottom.
- ***Good intentions but inability to follow through:*** e.g. stating an intention to donate to charity in good faith, but being unable to carry it out due to lack of resources (e.g. falling into debt), inability to implement it (e.g. not possessing the technical knowledge needed to

go about withdrawing money from one's account and donating it, etc.), or weakness of will (e.g. the difficulty encountered in actually trying to keep the intention).

- ***Balancing out good deeds with bad ones:*** e.g. committing to being healthy and exercising vigorously in the morning and then eating chocolate cake for lunch - such an individual believes they have achieved the desired goal of overall 'healthiness', as the unhealthy behaviour was balanced out by the healthy one.

It is also possible that such inconsistency can arise as a result of ambivalence about one's own desires and beliefs. From a certain point of view, such cases may look like integrity, as the individual may not have a sense of dis-integration. The question, then, is whether or not such an ambivalence to one's own desires and beliefs precludes the possibility of integrity. Consider the following:

- ***The case of two-culture children:*** e.g. the child who stays with her family and wholeheartedly decides not to resolve conflicts between the cultures, vs. the one who decides to run away – the runaway does not necessarily demonstrate greater integrity (Musschenga 2001).
- ***The case of oppressed groups:*** Oppressed groups may remain ambivalent between their understanding of their motives and actions, and the dominant understandings of the motives and actions given to them by the structures or groups oppressing them. This ambivalence is held out of a desire to hold competing perspectives on a kind of par or level playing field. Indeed, Calhoun (1995) relates how, "Ambivalence is endemic amongst oppressed groups who suspect that dominant interpretations of their motives and actions are mistaken but for whom there are as yet no clearly articulated arguments discrediting dominant views. [...] Agents can have reason to resist resolving ambivalence. In particular, they may think it important to acknowledge a basic assumption underlying practical deliberation, namely, the equality of deliberators."

Finally, there may be times in which we need to hold inconsistent beliefs or desires, in order to see how to best work them out in the long term (see MacIntyre 2008, although he only discusses desires).

- ***Desires Cases:*** Consider an individual who desires both to have a slim figure and also to eat whatever he wants. He may, in the short term, hold both desires (even though they are inconsistent with one another), in order to gain more information about how they work, how they conflict, and how they fit with his other desires - in order to be better placed to choose to commit to one of these desires (and maintain this commitment, and know how to best work it out and implement it) in the long term.
- ***Belief Cases:*** Consider an individual who wants to know whether Western medicine's chemotherapy or faith healing is more effective in treating his cancer. He may hold a belief in both simultaneously, as he tries to compare the beliefs in each system to see which has

more explanatory power to explain and affect the course of his cancer. Once one reveals itself to have more explanatory power and more positive effects upon the course of his cancer, he will discard the other.

An account of integrity, then, will need to be able to provide a ruling on whether these difficult test cases count as integrity or not. These test cases are so challenging because our intuitions seem to cut both ways with these cases; however, these conflicting intuitions provide useful data that we will leverage to show how our new model of integrity can make sense of our reactions to these cases. We will return to these cases in Section 3, when we will show how our new model is better able to explain these cases.

## 1.2. Conceptual Clarification

What is integrity? The definition of integrity is important, as some develop their definitions in order to distinguish between the problem cases above in terms of which do or don't count as integrity.

For example, does “integrity” only come in a moral form, pertaining only to issues of kindness/well-being and justice/fairness, or might it also come in an epistemic form, concerned with cognitive contact with reality (Wilson 2017)? Or perhaps integrity comes in a meta or “all-things-considered” form that balances both moral and epistemic considerations? Perhaps there are even further categories beyond the moral and epistemic, such as “professional integrity,” which may not be reducible to the moral or epistemic forms.

To see how the test cases are helpful to these conceptual questions, suppose that integrity comes only in a moral form. In such a case, we need not concern ourselves with some of the test cases involving holding conflicting beliefs (where those beliefs are epistemic and not moral). Additionally, the moral form of integrity may well preclude the possibility of possessing integrity due to one's consistent commitments to evil and harm (as in the Hitler and Milton's Satan cases). This discussion is just intended to show how the test cases can help to focus and chart the conceptual space around the concept of integrity.

Additionally, one way of differentiating the virtues is through looking at their characteristic behavioural, motivational, and affective structures. One might think that integrity's characteristic behaviour involves engaging in a kind of continual practice of reflective equilibrium to order one's beliefs, desires, identities, and commitments in a consistent way and ensure they cohere with one's behaviour. Alternatively, one might think that integrity may be one of the few virtues (or perhaps the only virtue) without a characteristic behavioral signature - perhaps integrity is defined by the absence of certain kinds of activity (i.e. the absence of conflict among some combination of one's beliefs, desires, commitments, behaviour, etc.) rather than by any positive behaviour. Regarding the motivational structure, it may be that integrity's motivational structure is the desire to achieve coherence and consistency across one's beliefs, desires, commitments, identities, and behaviours.

What might the affective structure be? Is it the case, as some psychologists have recently suggested, that joy is the result of experiencing a sense of integrity: integration between how the world is and how one hopes it might be (Van Cappellen 2020) or a sense of integration within the self (Johnson 2020b). Additionally, are there other emotions that might facilitate or support integrity, rather than resulting from integrity?

Furthermore, we may well ask whether integrity even is a virtue, or whether it is better thought of as the combination of all of the virtues, as a meta-virtue, or as a precondition for virtue or human functioning (Flanagan 1991, p. 10; Frankfurt 1988: 159-176, 2004: 95-98; see also Johnson 2020c: 177, fn 27). In asking these questions, it raises the further question as to how integrity is related to, yet distinct from, other virtues and related psychological constructs such as honesty, courage, authenticity, self-efficacy, etc. It is important to be able to map where these concepts overlap, correlate, and come apart.

Finally, we might ask what the vices of deficiency and excess are for integrity (if they do, in fact, exist) and what integrity's structural opposite is (if those are different). For instance, Taylor (1981) lists the following options for the structural opposite of integrity: (1) *moral hypocrisy* (only pretending to be committed), (2) *Shallow sincerity* (either not being committed to anything enough, or else being sincere about incompatible convictions), (3) *Weakness of will* (having commitments but not remaining true to them), (4) *Self-deception* ("self-deceptively taking oneself to be committed to some project"). Alternatively, and more straightforwardly, perhaps the structural opposite of integrity is simply fragmentation, chaos, or inconsistency in one's beliefs, desires, identities, commitments, and behaviours. What might be the vice of deficiency for integrity? It might also involve this kind of fragmentation, chaos, or lack of integration. And what of the vice of excess? It may be the case that there is no corresponding vice of excess (some virtues may not have a vice of excess or deficiency), or it might involve having too strong of a drive toward integration, precluding the ability to hold the kind of lack of integration that may be needed in some cases (e.g. in the aforementioned cases of oppression, or when one needs to hold conflicting desires or beliefs to see how they play out in the long term).

### **1.3. Empirical Investigations of Integrity**

There are a host of models of integrity that are in use, and this conceptual confusion regarding integrity leads to issues when it comes to empirically investigating it. For example, some see integrity as largely involving a kind of honesty, or as making good on one's commitments or aligning one's professed moral commitments with one's behaviour (i.e. it is the opposite of a kind of moral hypocrisy). Others see integrity as involving a kind of psychic unity (psychic integration): the goal of many forms of psychotherapy is to achieve psychic integration. A challenge, then, is whether integrity might be distinct enough from other virtues (e.g. honesty) and psychological constructs (e.g. psychic unity) to be able to test for it.

Furthermore, it is also important to investigate the correlates and predictors of integrity, as these may indicate the additional resources and skills that are needed in order to exercise integrity. For instance, “conscientiousness” is one of the strongest predictors of integrity (Murphy & Lee 1994): this may be because conscientiousness is required in order to carefully and constantly monitor and evaluate one’s beliefs, desires, commitments, and behaviour, in order to achieve coherence across all of them. Additionally, something like self-efficacy may be required in order to achieve integrity, such that one is able to bring about the changes in oneself (and through one’s behavior) to which one is committed.

A new model of integrity, then, is needed that can organise the existing empirical literature in this conceptual space and also yield testable new predictions. It is to such a model that we now turn.

## **2. A New Definition of Integrity**

We propose a new three-fold definition of integrity: integrity involves (1) the integration of self-to-world (whereby one’s relationship to the world is ordered by the way the world really is); (2) the integration of world-to-self (whereby one is able to change the world to cohere with one’s desires); and (3) internal integration (which involves psychic unity or integration within oneself). The first two forms of integrity build on the models of the intentional states of belief and desire developed by Searle (1995, 2001): indeed, both (1) and (2) have the same “conditions of satisfaction” (i.e. coherence or alignment between something in the self and something in the world); however, both have different “directions of fit.” In (1), one must change something within oneself (e.g. change a belief in one’s noetic structure, or a commitment or identity that one has) to “fit” with the way the world is; whereas (2) involves changing something in the world to “fit” with something within one’s mind (e.g. to cohere with one’s desires). While the terminology can be difficult to keep straight, we can think of (1) as “mind-must-conform to world” and (2) as “world-must-conform-to-mind” (see Søraker 2012, fn 4).

### **2.1. Integration of Self-to-World**

Firstly, there is the *integration of the self to the world*. This comes in an epistemic form, whereby one’s internal representations track the way the world is, and a moral form, whereby one’s commitments are ordered by the moral responsibilities that the world places on oneself. The epistemic form, then, involves avoiding self-deception (Taylor 1981), while the moral form involves care and “sensitivity to the needs of others” (Peterson & Seligman 2004, 250).

## **2.2. Integration of World-to-Self**

There is the *integration of the world to the self*, in which one's behaviour is able to bring the world into alignment with one's commitments and desires. This form, then, involves related constructs such as self-efficacy (the ability to successfully act on one's desires) and self-integrity (the ability to maintain one's sense of oneself as good and moral, through one's actions; Cohen & Sherman 2014). This can come in a specifically moral form, in which that behaviour is consistent with one's moral commitments, in particular (Monin & Merritt 2012; Calhoun 1995; Korsgaard 2009).

## **2.3. Internal Integration**

Thirdly, there is *internal integration*. This comes in an epistemic form (in which all of one's beliefs in one's noetic structure cohere), a moral form (in which one's commitments cohere), and an overall form that sits above the two (in which all of one's beliefs, desires, commitments, and identities cohere). Internal integration, then, involves a kind of psychic unity or psychic integration, which has traditionally been seen as a precondition of having virtue (Flanagan 1991, p. 10). Internal integration also involves a kind of "wholeheartedness" (Frankfurt 1988: 159-176, 2004: 95-98) characterised by absence of internal conflict over one's beliefs, desires, commitments, and identities. Such psychic unity is required for flourishing, as without it, we are at risk of alienation from the self or a "drastic psychic injury" (1998: 139).

# **3. Advantages of the New Definition**

The new, three-fold definition, can help to clear up the conceptual confusion and more clearly map the conceptual space around the virtue. For example, it shows how correlates or predictors of the virtue relate: honesty/avoiding self-deception and "sensitivity" are involved in the integration of self-to-world, self-efficacy and self-integrity are involved in the integration of world-to-self, psychic unity and wholeheartedness are involved in internal integration, etc.

The new definition is able to also explain how and why a variety of related psychological phenomena occur. For example, the literatures on self-verification theory and stereotype threat indicate that individuals will act in congruence with an identity that has been externally ascribed to them, even if the identity is deviant or undesirable (see e.g. Kraut 1973, Lemert 1972, Swan & Ely 2004, Dotter & Roebuck 1988; see also Vellman 2005, Alfano 2013, Johnson 2020c for reviews). Such behaviours can be understood as an attempt to reduce cognitive dissonance, conflict, and uncertainty within one's beliefs, desires, commitments, identities, and behavior. This happens through adjusting (or confirming) beliefs about oneself (self-to-world), adjusting one's behaviours (world-to-self), and re-ordering one's beliefs, desires, commitments, and identities (internal



integration), so that these all cohere with the identities that have been externally ascribed to oneself.

This three-fold approach can also sort through the morality - epistemic interaction, showing how the various forms can come in both moral and epistemic forms, and how they moral and epistemic forms can be brought into a greater level of coherence through a higher-order form of integrity.

We turn now to the test cases with which we started this paper, in which our intuitions cut both ways. The data of these conflicting intuitions is able to be explained by the three-fold model, providing further support for this model.

- ***Being fully committed to (and consistent in) doing harm.*** Such cases (e.g. Hitler, Milton's Satan, etc.) do not involve integrity, as they fail in the moral form of the integration of self-to-world: such a commitment to harm or evil involves a lack of sensitivity to the moral responsibilities that the world places on the self.
- ***Doing harm in the course of being committed to some project.*** Such cases (e.g. Gauguin leaving his family to pursue his art) do not involve integrity, as they likewise fail in the moral form of the integration of self-to-world.
- ***Making a choice when faced with a moral dilemma.*** Truly intractable moral dilemmas threaten integrity as one experiences a pull to both, and is therefore not "wholehearted." Furthermore, after having made the decision one may experience a sense of "moral remainder" that also threatens psychic unity (this will be especially likely in the case of "tragic dilemmas"). In order to achieve integrity, one must either fully give up a commitment to one of the two options (to thereby achieve "wholeheartedness") and likewise give up their sense of "moral remainder" (to thereby achieve psychic unity). How this works becomes clearer with consideration of the Williams' cases:
- ***Cases from Bernard Williams***
  - Jim needs to be able to either give up his commitment to *not harming the one*, thereby making it possible to wholeheartedly shoot the one individual, or give up his commitment to *saving the twenty*, thereby making it possible to wholeheartedly restrain from intervening (by shooting the one) as the twenty are slain. Additionally, Jim must then, after the point of decision, give up his sense of moral remainder in order to achieve psychic unity. Importantly, achieving wholeheartedness or giving up the sense of moral remainder may well also preclude the possibility of integrity: it may be that there are still demands upon Jim (e.g. to not shoot an innocent person, to save the twenty) that cannot be waved away simply because the situation is intractable (this is what happens in Tessman's account of 'tragic dilemmas,' if such cases are indeed possible in which 'doing the right thing is impossible'). Wholeheartedly committing to one or the other courses of action, then, would involve a failure of integration of self-to-world, by rendering oneself insensitive to some moral responsibility that the world places on oneself. Similarly, giving up one's sense of moral remainder may also involve a failure of integration of self-to-world, by rendering oneself insensitive to some aspect of the moral fabric of the world to which one

ought not be insensitive (an aspect to which one is properly responsive if one experiences such a sense of moral remainder).

- ***Acts committed in situations of systemic injustice.*** In such cases, certain integrations of self-to-world (following laws) are overridden by higher-order moral responsibilities in the world (to justice, liberation, etc.). One can, therefore, be wholehearted in one's realisation that it is situationally appropriate to reject the lower-order moral responsibilities in such a case.

The cases in which there may be integrity despite inconsistency or lack of commitment:

- ***Preaching without practicing:*** such cases involve a failure of integration of self-to-world. For instance, the smoking doctor fails to be properly responsive to a moral responsibility the world places on him (to take care of his health).
- ***Not practicing and then preaching:*** such cases involve true integration, provided that one's future behavior is consistent with the new identities, desires, and commitments that one has adopted.
- ***Good intentions but inability to follow through:*** such cases involve a failure of integrity of world-to-self in that they involve failures of self-efficacy.
- ***Balancing out good deeds with bad ones:*** such cases involve failures of proximal integrity, while they may involve achieving ultimate integrity. For instance, the person who exercises vigorously in the morning and rewards themselves with a slice of chocolate cake for lunch fails in that moment to act with integrity in the proximal sense; however, they are succeeding in a more ultimate pursuit of integrity, particularly if eating the cake helps them to maintain the healthy lifestyle overall (c.f. how periodic "cheat days" of junk food binges function as rewards that sustain the broader disciplined eating patterns of bodybuilders). This, then, helps to provide an account of how the psychological phenomenon of 'moral licensing' may work, and how it relates to integrity. There is, however, a higher level of integrity achieved if one's proximal and ultimate integrity are aligned.

It is also possible that such inconsistency can arise as a result of ambivalence about one's own desires and beliefs:

- ***The cases of two-culture children and oppressed groups:*** At stake are integration of self-to-world (being responsive to the moral responsibilities that the world places upon oneself) and internal integration (psychic unity). For example, the child who stays with her family and wholeheartedly decides not to resolve conflicts between the cultures may achieve integration of self-to-world in some forms (by honouring the responsibilities they have to their family) while threatening the integration of self-to-world in some other forms (by not honouring responsibilities they may have to themselves, to avoid self-alienation). Furthermore, they may risk psychic unity, by holding contradictory identities, beliefs, and commitments due to the two cultures. Compare them to the individual who runs away: they may avoid self-alienation, but they fail in integration of self-to-world, by not honouring

the responsibilities they may have to their family. Such ambivalence about one's responsibilities involves a failure of integration of self-to-world, by not being appropriately responsive to the demands the world places upon one, showing that attempting to achieve integrity through cultivating ambivalence about one's desires does not actually conduce to full integrity (the same can be said for beliefs, as it also involves a failure to be appropriately responsive to the epistemic demands the world places on oneself regarding the integration of self-to-world). Faced with either staying (and risking self-alienation) or leaving (and failing in the integration of self-to-world), such individuals may be placed in 'tragic dilemmas' where full integrity is not possible. Indeed, Grace Ji-Sun Kim, in a discussion of how Asian Americans are caught between conflicting Asian and American cultures, relates how: "Marginal people's self-alienation stems from their external alienation by dominant societies. The self is split in two when marginalised people are torn between two worlds. Sociologist Everett Stonequist states, 'The duality of cultures produces a duality of personality - a divided self.'" (Kim 2001: 66).

Finally, there may be times in which we need to hold inconsistent beliefs or desires, in order to see how to best work them out in the long term. Such cases are best conceptualised through our recently introduced distinction between proximal and ultimate integrity: here, proximal integrity is sacrificed in the service of achieving ultimate integrity.

A final world on these test cases is in order, which is that they indicate the correlation of integrity and *phronesis*, or practical wisdom. Indeed, for cases in which proximal integrity must be sacrificed in the pursuit of ultimate integrity, or in cases in which we encounter dilemmas due to conflicting identities, commitments, and desires, we will need *phronesis* to help determine which of these competing actions, identities, commitments, and desires to choose in that particular situation. A major function of practical wisdom, then, is to harmonise our moral perceptions, commitments, and actions, both across our lives and within our psyches, enabling us to achieve overall integrity (see also Kamtekar 2004).

## **4. Structural Conditions for Integrity**

We turn now to the structural conditions needed for integrity, in order to provide insights for practitioners hoping to foster and cultivate integrity.

Firstly, integrity may allow for cross-situational consistency, and in this way lessen the force of the situationist critique. Failures of consistent behavior, which appear at first blush to provide evidence for situationism, might then be better understood as cases in which one has failed in conscientiousness (i.e. failed to align commitments and behavior), and therefore in integrity. By

cultivating integrity and its predictors, one may then achieve integration across one's commitments and behavior.

Secondly, there may still be structural constraints on the development of integrity in some situations. Here, we extend Babbitt's (1997: 118) suggestion that, "some social structures are of the wrong sort altogether for some individuals to be able to pursue personal integrity."

#### **4.1. Babbitt on the interaction between social structures and integrity**

Babbitt (1996: 105) makes a distinction between *individuation* and *adequate* identity. *Individuation* refers to how people are perceived as individuals, which includes how others perceive them and also how they are aware of themselves. *Adequate* identity is a normative notion, referring to a self-concept which forms an "adequate basis for deliberation and action," bringing about genuine individuality.

Given this distinction, it is possible that individuation and adequate identity can come apart, and so reaching an 'adequate' sense of identity takes work of discovery and development. This is especially problematic in situations of systemic injustice, in which traits (e.g. perceived race) become identifying "not primarily as a result of individual character but because society possesses a certain character." For example, in America, due to various societal schemas and constraints, it is often more difficult for Black women to be understood in terms of aspects that actually explain their choices and actions. In such situations, it can be difficult for someone to reach an adequate sense of their own individuality. On account of these structural constraints and *hermeneutical injustices*, they may not have the epistemic resources to sufficiently examine their own beliefs and principles (and the cognitive load of constant marginalisation and oppression can diminish their ability to exercise these resources), and to identify the choices available to them. If the "self" to which one is supposed to be true remains undefined, this makes it difficult even to raise the question of whether one is acting reasonably or whether one can be "true" to oneself or act in one's own best interests (i.e. for one's own dignity and self-worth).

Consequently, personal integrity is not just about facts about the individual (e.g. their interests, choices, beliefs and principles); nor is integrity solely about how facts about the individual fit within their society. Social structures already determine which properties are considered to be identifying, and which epistemic resources are available. "Pursuing personal integrity may be less about understanding one's identity and beliefs, and more about what one's society is and what it could be" (Babbitt 1997: 118).

Babbitt backs up this point by discussing cases in which people may look "crazy" and "self-destructive," when they act in ways which disrupt their existing fundamental commitments (and those of society). In some of these cases, people can be shown to have acted with integrity once

the wider context of systemic injustice in which they acted is considered. Indeed, such an act can be understandable when it is seen in terms of its role in developing a different kind of society, which often involves disrupting rather than upholding the fundamental values of an unjust society.

Babbitt gives the example of conscientious objectors during the Vietnam War era. She suggests that they are more accurately described as “losing” (rather than “finding”) their own identity and beliefs. The discovery of what their society was like, and the kind of structures that determined their identity (“greed” and “imperialism”), motivated them to lose that identity.

The impact of social structures is relevant not only for the oppressed but also for those who benefit from such structures. The latter group will also be constrained to understand themselves in terms of unjust structures. For both groups, reaching an adequate sense of self will take work.

- Compare Calhoun: “I think it is important to be skeptical about any account of integrity whose implication is that members of oppressed groups are particularly likely not to have integrity or that, for them, acting with integrity requires acting in morally unsavory ways (for example, ignoring all but one oppressive system or adopting a dismissive stance toward social judgments). [...A]chieving the ideal of an integrated self does not depend solely on an agent's internal capacities. It also depends on social conditions. The illusion that integration is entirely up to the individual may reflect a particularly privileged social position – for example, one from which the question of where one stands with respect to multiple and conflicting oppressions does not regularly come up and within which one's own self-interpretation receives substantial social confirmation.” (Calhoun 1995: 241, fn. 12)
- Also compare Peterson & Seligman (2004: 270): “[O]ne can imagine a person who in her behavior is true to a false concept of herself, while being false to her true personality (i.e., an extroverted woman fails to express her beliefs because her culture trains her to think that women have nothing to say). Indeed, a double negative may apply, in that in some cases being false to one's false self-concept may be a route to heightened growth and adaptation!”

Consequently, social structures will provide the epistemic constraints that may go a long way to determining the integrity available to both those marginalised by, and those who benefit from, these structures. Furthermore, apparent cases of failures of integrity can sometimes instead be seen as failures of proximal integrity in service of ultimate integrity, once one understands them in the context of attempting to change those unjust social structures to enable the possibility of achievement of more ultimate forms of integrity (for oneself or others) within that society.

## **4.2. Identifying the different structural conditions using our framework**

Our three-fold framework reveals ways in which structural conditions can impede the development of integrity. Babbitt focuses on the impact of structural injustice on one kind of integrity (integration of self-to-world), but our broader definition of integrity allows us to outline constraints on additional forms of integrity. As just discussed, integration of self-to-world is threatened by social structures that impede one's epistemic resources: here one is unable to rightly order one's beliefs and identities to the way the world is, or to be appropriately sensitive to the moral responsibilities the world places upon oneself. Additionally, integration of world-to-self is threatened by social structures that impede self-efficacy, as one will not be able to accord their actions with their desires and commitments. Finally, internal integration is threatened by social structures that promote self-alienation, as one's identities, beliefs, desires, commitments, and actions cannot cohere. Such individuals may often find themselves in 'tragic dilemmas.'

## **4.3. Suggestions for Overcoming Constraints**

Babbitt highlights how constraints on integrity can be epistemic in nature. Individuals may not have the conceptual resources to find out what social structures make them what they are and to formulate and pursue an adequate self-concept. How can such constraints be overcome?

The challenge is that merely providing spaces in which people are encouraged to ask questions and critically reflect (for a recommendation for how social workers may develop professional integrity through space for critical reflection, see Banks 2016) may not be enough, if they are not also provided with the right conceptual resources.

Indeed, Medina's (2013) work on 'meta-lucidity' shows how, paradoxically, those in certain situations of structural oppression are often best situated to identify their constraints and re-order their conflicting commitments to achieve integrity. Living through injustice can heighten awareness of the disconnect between how society individuates you and what really motivates and informs your beliefs and actions.

However, how available is 'meta-lucidity' to those who benefit from oppressive social structures? It seems to require the motivation for personal change. Achieving integrity in this way, therefore, may well not primarily be about availability of information, but rather the appeal (or lack thereof) of personal change. For both the oppressed and those who benefit from social structures, the primary constraint on the development of integrity may not be epistemic (the availability of information and resources) but the lack of motivation to pursue personal change of the sort that might involve vulnerability and resistance. Such a motivation may well be difficult to achieve, as the state of possessing doubt about self-knowledge, or lacking a stable self-identity may be highly aversive (Vellman 2000).

A way to tackle this may be to develop a more collective understanding of integrity. We may need to develop our imagination of what it is for structures to be good for everyone, making it plausible that until these structures are developed, integrity is impaired for all people. We may also need to highlight the importance of developing solidarity (Banks, 2016). The usual image of integrity may be of the isolated individual standing alone for their beliefs, but the development of integrity may be better fostered by being part of a supportive group. This is consistent with both Aristotelian and Confucian virtue theoretical traditions that emphasise the importance of friendship and virtue formed through membership in a community of individuals jointly cultivating virtue. Indeed, rather than the model of virtues of atomised individuals developed in isolation, it may well be that only by being in the right relationships of friendship, where groups try to develop common virtues, that such integrity is possible. Such a move is also consonant with recent approaches to ethics that shift away from models that involve dyads of perpetrators and victims or dyads of privileged and oppressed, and toward models of co-liberation (see e.g. D'Ignazio & Klein 2020). 'Meta-lucidity,' then, is best developed in the context of community, friendship, and co-liberation: without such solidarity, one will lack the motivation to pursue the personal change needed to achieve more ultimate forms of integrity.

## 5. Conclusion

Current approaches to integrity fail to sufficiently capture the conceptual space around the virtue and explain difficult test cases. In this paper, we have shown how our novel, three-fold model of integrity is better able to map this conceptual space - illustrating how integrity overlaps with some concepts (such as honesty) and correlates with other, related concepts (such as conscientiousness). We have also shown how our model is able to make sense of our conflicting intuitions in these integrity test cases. As further empirical work on integrity and related constructs and capacities emerges, our model will receive even further support if it is able to accommodate these additional developments in the literature. In the second part of the paper, we turned to implications of this model for practitioners, by looking at how social structures can provide constraints on the three forms of integrity (for both oppressed and privileged groups), and how they can be overcome. We have here suggested the importance of the cultivation of the capacity for 'meta-lucidity' (Medina 2013) and of shifting from an emphasis on the cultivation of individual virtues in atomised individuals, to the cultivation of virtues of collectives in community and frameworks of co-liberation.

## References

- Alfano, M. (2013). *Character as moral fiction*. Cambridge: Cambridge University Press.
- Babbitt, Susan E., 1997. 'Personal Integrity, Politics and Moral Imagination,' in S. Brennan, T. Banks, S.(2016). "Professional Integrity: From Conformity to Commitment." in R. Hugman & J. Carter (eds.), *Rethinking Values and Ethics in Social Work* (pp. 49-63). London: Palgrave Macmillan. T. L. Isaacs and M. Milde (eds.), *A Question of Values: New Canadian Perspectives on Ethics and Political Philosophy* (pp. 107-31), Amsterdam and Atlanta: Rodopi.
- Calhoun, C. (1995). Standing for Something. *Journal of Philosophy*, 92, 235–260.
- Cohen, G. L. & Sherman, D. K. (2014). The Psychology of Change: Self-Affirmation and Social Psychological Intervention. *Annual Review of Psychology*, 65(1), 333-371.
- D'Ignazio, C. D. & Klein, L. F. (2020). *Data feminism*. Cambridge, MA: MIT Press.
- Dotter, D., and Roebuck, J. (1988). The labeling approach re-examined: interactionism and the components of deviance. *Deviant Behavior*, 9(19), 19-32.
- Flanagan, O. (1991). *Varieties of Moral Personality*. Cambridge, MA: Harvard University Press.
- Frankfurt, H. (1987). "Identification and wholeheartedness," in F. Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (pp. 27-45). New York: Cambridge University Press.
- Frankfurt, H. (1988). *The importance of what we care about: Philosophical essays*. Cambridge: Cambridge University Press.
- Frankfurt, H. (2004). *The reasons of love*. Princeton: Princeton University Press.
- Johnson, M. K. (2020a). Joy: A reply to the replies. *Journal of Positive Psychology*, 15(1), 84-88.
- Johnson, M. K. (2020b). Joy: A review of the literature and suggestions for future directions. *Journal of Positive Psychology*, 15(1), 5-24.
- Johnson, M. K. (2020c). *Imaginative simulation in the moral life*. Doctoral Dissertation. University of Cambridge.
- Kamtekar, R. (2004). Situationism and virtue ethics on the content of our character. *Ethics*, 114(3), 458-491.
- Kim, G. J-S. (2021). *Invisible: Theology and the experience of Asian American women*. Minneapolis: Fortress Press.
- Korsgaard, C. (2009). *Self-Constitution: Agency, identity, and integrity*. Oxford: Oxford University Press.
- Kraut, R. (1973). Effects of social labeling on giving to charity. *Journal of Experimental Social Psychology*, 9(6), 551 –62.
- Lemert, E. (1972). *Human Deviance, Social Problems and Social Control*. New York: Prentice Hall.
- MacIntyre, A. (2008). "Conflicts of desire." In J. P. Dougherty, & T. Hoffman (Eds.), *Weakness of Will from Plato to the Present* (Studies in Philosophy and the History of Philosophy, Vol. 49, pp. 276-292). Washington, DC: Catholic University of America Press.
- Medina, J. (2013). *The epistemology of resistance: Gender and racial oppression, epistemic injustice, and the social imagination*. Oxford: Oxford University Press.
- Milton, J. (1667/2000). *Paradise lost*. New York, NY: Penguin Books.
- Monin, B., & Merritt, A. (2012). "Moral hypocrisy, moral inconsistency, and the struggle for



- moral integrity.” In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 167–184). American Psychological Association.
- Murphy, K.R., Lee, S.L. (1994). Personality variables related to integrity test scores: The role of conscientiousness. *Journal of Business Psychology*, 8, 413–424.
- Musschenga, A. (2001). Education for moral integrity. *Journal of Philosophy of Education*, 35(2), 219–235.
- Palanski, M. E., & Yammarino, F. J. (2007). Integrity and Leadership: Clearing the Conceptual Confusion. *European Management Journal*, 25(3), 171–184.
- Peterson, M. & Seligman, M. (2004). *Character strengths and virtues: A handbook and classification*. Oxford: Oxford University Press.
- Searle, J. (1995). *The construction of social reality*. New York: Free Press.
- Searle, J. (2001). *Rationality in Action: The Jean Nicod Lectures*. Cambridge, Mass.: MIT Press.
- Søraker, J. (2012). Virtual Worlds and their Challenge to Philosophy: Understanding the ‘Intravirtual’ and the ‘Extravirtual’. *Metaphilosophy*, 43(4), 499-512.
- Swan, W. B., & Ely, R. J. (1984). A battle of wills: Self-verification versus behavioral confirmation. *Journal of Personality and Social Psychology*, 46(6), 1287-1302.
- Tessman, L. (2015). *Moral failure: On the impossible demands of morality*. Oxford: Oxford University Press.
- Tessman, L. (2017). *When doing the right thing is impossible*. New York NY: Oxford University Press.
- Taylor, G. (1981). Integrity. *Proceedings of the Aristotelian Society*, Sup. Vol. 55, 143-159.
- Van Cappellen, P. (2020). The emotion of joy: Commentary on Johnson. *Journal of Positive Psychology*, 15(1), 40-43.
- Velleman, D. (2005). *Self to Self: Selected Essays*. Cambridge: Cambridge University Press.
- Williams, B. (1973). “A Critique of Utilitarianism.” In J. J. C. Smart & B. Williams, B. (Eds.), *Utilitarianism: for and against* (pp. 75-150). Cambridge: Cambridge University Press.
- Williams, B. (1981). *Moral luck*. Cambridge: Cambridge University Press.
- Wilson, A. T. (2017). Avoiding the Conflation of Moral and Intellectual Virtues. *Ethical Theory and Moral Practice*, 20(5), 1037–1050.