FILOZOFIA
Volume 81, Number 3, 2026

DOI: https://doi.org/10.31577/filozofia.2026.81.3.1

Freedom and Commitment: Does Kant Hold a Subjectivist Theory of Freedom?

GALEN STRAWSON, Department of Philosophy, University of Texas at Austin, USA

STRAWSON, G.: Freedom and Commitment: Does Kant Hold a Subjectivist Theory of Freedom?
FILOZOFIA, 81, 2026, No 3, pp. 277 – 293

According to a *Subjectivist* theory of X, reference to belief in the existence of X enters essentially into a full account of what the existence of X consists in. A Subjectivist theory doesn't claim that X is a 'merely subjective' phenomenon; it's a theory of what the existence of X consists in as a matter of objective fact. A *Commitment* theory of X is a theory of X that holds that we are inescapably committed to believing in the existence of X. It may also be a Subjectivist theory of X and hold that our commitment to belief in the existence of X is part of what makes it the case that X exists. This paper aims to specify the character of Subjectivist and Commitment theories, before suggesting that Kant may be said to have held a Subjectivist and indeed Commitment theory of freedom.

Keywords: Kant – free will – commitment theory of freedom – moral law – cognitive phenomenology – self-creation

I. Introduction: Subjectivist Theories

A *Subjectivist* theory of X is a theory according to which reference to belief in the existence of X enters essentially into a full account of what the existence of X consists in. It doesn't involve the claim that X is a 'merely subjective' phenomenon: a Subjectivist theory of X is a theory of what the existence of X consists in as a matter of objective fact – if it exists at all.

It's natural to think that there can never be a true Subjectivist theory of anything that exists, because the existence of a belief can never be part of what makes that same belief true. One might call this the *Principle of Independence*. Spelt out a little, it states that if one has a belief, B, then the obtaining of the state of affairs that makes B true cannot (ever) depend upon or necessarily

involve one's having B. One's having B cannot be among the truth-conditions of B; a true belief must always be supposed to be a representation of some state of affairs essentially other than itself. It must be something which can always in principle be subtracted from the world in a way that leaves its object – the state of affairs it is a belief about – untouched. So how could there be a true Subjectivist theory of anything?

Perhaps the only cases in which Subjectivist theories have even *prima facie* plausibility are cases that involve beliefs about a conscious or experiencing being, where the belief in question is a belief on the part of that same conscious being. According to a Subjectivist theory of fashionableness, for example, Lucy is fashionable only if she believes herself to be fashionable.[1] According to a Subjectivist theory of promising, Lucinda makes a promise at time *t* only if she believes at *t* that she is making a promise.[2]

One can replace the word 'believe' by other terms when assessing the prospects for Subjectivist theories. It's arguable that Louise enters into a contract, at a given time *t*, only if she believes at *t* that she is entering into a contract, but one can put this equally well by saying that she enters into a contract at *t* only if she conceives of herself at *t* as entering into a contract.[3] So too one can say that Louis is truly fashionable only if he *experiences* or *takes* himself to be fashionable.

Consider the artificial non-factive term 'is aware*', which stands to the factive term 'is aware' more or less as the non-factive term 'believes' stands to the factive term 'knows'. Given this new term one can say that Ludwig enters into a contract, at *t*, only if he is at *t* aware* of himself as entering into a contract. And if the Subjectivist theory of entering into a contract is correct one can drop the asterisk and use the normal term 'is aware' to observe that Ludwig enters into a contract or makes a promise at *t* only if he is at *t* aware that he is entering into a contract or making a promise. It seems that there's an *awareness condition* on such actions.

The general idea is simple. Some properties ϕ are such that experiencing oneself as being ϕ (believing oneself to be ϕ) is a necessary condition of one's

[1] A woman from an isolated traditional society who arrives in Paris in her normal clothes and turns out to be dressed according to the latest fashion is not fashionable.

[2] Note that the belief must be about oneself conceived of specifically as oneself. For an account of this distinctive way of thinking of oneself see, e.g., Castañeda (1966), Perry (1979).

[3] This is not true in UK or US law, in fact, but it could be true in some jurisdiction.

being ϕ.[4] It's arguable that one acts immorally only if (or if) one takes oneself to be acting immorally. There is something paradoxical about some of these claims, but they can also seem eminently defensible.

To hold a Subjectivist theory of freedom – free agency – is to hold that believing one is a free agent – experiencing oneself as a free agent – is a necessary and indeed constitutive condition of actually being a free agent. A Subjectivist theory of moral responsibility states that conceiving of oneself as a morally responsible agent – taking oneself to be or grasping* oneself as a morally responsible agent (where 'grasp*' is non-factive in the same way as 'aware* that') is a necessary and indeed constitutive condition of actually being a morally responsible agent. Some may think that there's a similar condition on being a person.

II. Commitment Theories

A *Commitment* theory of freedom or moral responsibility[5] opens with the claim that we are inescapably committed to believing we're free agents, inescapably committed to believing we're radically free to choose in such a way that we can be held to be responsible for our actions in some very strong – absolute – sense. But it doesn't just hold that it's an extremely important natural fact about us that we can't help believing we're free agents – a view that is compatible with holding that we're not really free agents at all. It also holds that if we really are free agents then we are so essentially partly in virtue of the fact that we believe we are. It is therefore a Subjectivist theory of freedom.[6]

To hold that believing one is free is a necessary condition of being free is still compatible with holding that we aren't really free agents at all. A full blown Commitment theory of freedom goes further. It doesn't just hold that believing one is free is a necessary condition of being free. It also holds that

[4] Imagine a prize draw at a fête in which the condition of winning a prize is having a prime number ticket and believing one has won a prize. At the end of the day, the organizer calls everyone together (they're an ingenuous crowd, and one or two of them are optimists) and asks, 'Who believes they have won a prize?' Consider also Ludwig as he consciously entertains the thought content 'This (very) thought is puzzling' and experiences or believes it to be true, while Ludovic entertains the same content and doesn't experience or believe it to be true.

[5] I focus on freedom, but everything I say should apply equally to moral responsibility.

[6] Subjectivist theorists of freedom needn't be Commitment theorists, because they can hold that believing one is free is a necessary condition of being free without having any views about whether or not we can help believing ourselves to be (or experiencing ourselves as) free.

freedom is possible, indeed actual, a fully objectively existing phenomenon – in spite of the fact that believing one is free is a condition of being free.[7]

In what follows I want to spell out the characteristic claims of commitment approaches to freedom in a little more detail, and suggest that Kant attempts a Subjectivist Commitment theory of freedom, although he often takes himself to be attempting the merely negative task of showing that freedom is not a contradictory concept given deterministic causality in the phenomenal or empirical realm (see, e.g., Kant [1785] 1996, 102 – 106, Ak.4.456 – 460; Kant [1788] 1996, 177, Ak.5.46; Kant [1781 – 1787] 1933, A557 – 558 / B585 – 586).

III. A Commitment Theory of Freedom (I)

As I conceive it, a Commitment theory of freedom has three main stages. It may first seek to show

> (1) why belief in freedom is inevitable given certain forms of experience – where the forms of experience are specified without reference to the fact that they give rise to belief in freedom.

Belief in freedom may be argued to be inevitable for a being that experiences itself as something possessed of a power of choice, or a will, that is in some fundamental way independent of, i.e. not merely a determinate if complex function of, its particular pro-attitudes or motives generally considered. It may be argued to be inevitable for a being that experiences itself, considered as decider and deliberator, as something which is in some way essentially over and above all its particular motives, and which is therefore capable of making decisions independently of, and in the light of, those motives. Such a picture of the self may not be a true picture of anything that could actually exist, but it may be a true picture of how a being can seem to itself to be. After all, most human beings see themselves in this way, more or less explicitly. And the fact that this is so is crucial to the human phenomenology of freedom.

P. F. Strawson has argued that our commitment to belief in human freedom and moral responsibility is most fundamentally grounded in our natural 'reactive attitudes' to others, e.g., our attitudes of gratitude and resentment (see, e.g., Strawson 1962). This proposal falls under (1) above. An

[7] The claim that we are free agents only if we believe we are can also perhaps be understood in a communitarian way, as the claim that if we-plural (we human beings, we in the community) are properly called free agents, then this is essentially partly because we as a plurality believe we are. Here, however, I'm concerned with the claim that each individual is a free agent only if it believes of itself that it is a free agent.

alternative view, more in line with the proposal in the last paragraph, is that our commitment to belief in our radical freedom and moral responsibility for actions lies most fundamentally in our experience of our own actions.[8] This proposal also falls under (1) above.

In its second stage, a Commitment theory may seek to show

> (2) that human beings do standardly have, and perhaps inevitably have, a form of experience of the kind indicated in (1).

From (1) and (2) a Commitment theory can deduce that we (standardly or inevitably) believe that we are free.

So far the theory is merely a theory about commitment to belief in freedom, a theory about the causes of such commitment. It becomes a positive theory of freedom in moving on in its third stage to claim that

> (3) we are free, and we are so essentially partly because we not only believe we are free but also must believe (cannot help believing) we are free.

It doesn't of course claim that belief in freedom is sufficient for freedom, only that it's necessary, and sufficient given the fulfilment of certain other conditions (see (iv) below).

A fourth part of a full Commitment theory might seek to show in detail why

> (4) we have, and perhaps must have,[9] a form of experience of the kind specified in (1).

The defense of (4) may have both phylogenetic and ontogenetic aspects, drawing both on the theory of evolution (evolutionary psychology) and individual developmental psychology.

---

[8] See Section IX below and Strawson (2014). (When I cite a work, I give the date of first publication, or sometimes the date of composition, while any page reference is to the edition listed in the bibliography. In the case of quotations from languages other than English I give a reference to a standard translation but do not always use that translation. References to Kant's *Critique of Pure Reason* are in the standard A/B form. References of the form 'Ak.5.30' are to the Akademie edition of Kant's works.)

[9] It needn't be claimed that all (basically normal) human beings must have a form of experience which is such that they must believe they are free. The claim need only be that, having such and such a form of experience, people must believe they are free; and that they do in fact have such a form of experience.

IV. A Commitment Theory of Freedom (II)

A commitment-theoretic argument for freedom might accordingly proceed as follows:

(i) Every being that has F-type experience of self believes that it is a free agent (cf. (1)).

(ii) Lucy has F-type experience of self (cf. (2)).

Therefore

(iii) Lucy believes that she is a free agent [from (i), (ii)].

(iv) Lucy is G

where G is a specification of a set of further properties necessary but not sufficient for being a free agent, e.g., the properties of being a purposive agent and being self-conscious.

(v) Every being that believes it is free, and is G, is indeed free (cf. (3)).

Therefore

(vi) Lucy is a free agent [from (iii), (iv), (v)].

A commitment-theoretic rider might be added:

(vii) Every being that is of kind H has, or comes to have, F-type experience of self (cf. (4)).

where H remains to be specified (one suggestion is that H = G).

This may not look very promising, as an account of radical freedom. (3), and so (v), are fundamentally paradoxical. But then – I submit – nor does anything else.

It's worth noting that Commitment theorists are not 'as if' theorists. 'As if' theorists may make a variety of claims. They may claim that whatever the case may be with regard to more conventional theories of freedom, or Commitment theories of freedom, we would do well to live our lives *as if* we were free. Or their claim may be the Kantian-sounding claim that to live our lives as if we were free is a practical or moral necessity – a necessary supposition of pure moral-practical (as opposed to purely theoretico-speculative) reason. But the 'as if' clause suggests the possibility of exercising choice about the adoption of belief in freedom as a regulative principle for one's conduct, and the Subjectivist Commitment theorists – whose greatest regret may be that they

have not been able to find a satisfactory non-Subjectivist theory – are likely not only to deny that there can be any such question of choice, but also to react with Sidgwickian contempt to the (to them) facile compromising of the 'as if' theorists:

> I cannot fall back on the resource of thinking of myself under a moral necessity to regard all my duties *as if they were* commandments of God, although not entitled to hold speculatively that any such Supreme Being exists. I am so far from feeling bound to believe for purposes of practice what I see no ground for holding as a speculative truth, that I cannot even conceive the state of mind which these words seem to describe, except as a momentary half-wilful irrationality, committed in a violent access of philosophic despair (Sidgwick 1907, 507).

V. 'Neutral' and 'Rational' Freedom: Kant

According to Kant 'freedom is a mere *idea*, the objective reality of which can in no way be shown according to laws of nature' ([1785] 1996, 105, Ak.4.459). One who says this gives us reason to suppose that any positive theory of freedom that he may put forward will be of a special character; especially when he says such things, close to the incompatibilist's heart, as 'what we wish to understand, and will never understand, is how *predeterminism*, according to which voluntary [*willkürliche*] actions, as events, have their determining grounds in antecedent time, can he consistent with freedom' ([1793] 1960, 45 n., Ak.6.49 – 50 n.).

As is well known, Kant sought to ground the possibility of freedom of the will in the unknowable noumenal self, in a way that showed freedom to be compatible with the holding of (deterministic) laws of nature in what he called the 'empirically real' world (see, e.g., Rosen 1988 – 1989). Here I want to suggest that Kant also tries to establish our freedom in another way, a way connected to but distinguishable from his argument to freedom from the fact of our consciousness of the moral law. In particular, I want to argue that he advances a Commitment theory of freedom, holding not only that we really are free in the ordinary strong sense, but also that this is partly because, as rational (and self-conscious) beings, we can't help regarding ourselves as free.

Such a suggestion is not only paradoxical; it also conflicts with some of Kant's dauntingly complex views. Certain caveats and restrictions are called for. The first is this: I will not go into the question of what exactly 'we' really are, on Kant's view – noumenal selves. For the most part I'll take it that we are what we think we are, human beings with real physical bodies.

Secondly, I'll be concerned with Kant's theory of freedom only in so far as it is a theory of freedom to choose, a theory according to which *a choice of the*

morally worse of two options may be a truly free choice. I will not be concerned with Kant's other conception of freedom, according to which a rational agent's action is truly free only if it is in accordance with reason, or Reason, or the Moral Law; for according to this second conception, a truly free choice is always a morally correct choice.

At certain points Kant tried to combine these two lines of thought, but they are for the most part pretty clearly distinguishable, and my present concern – in Sidgwick's helpful terms (Sidgwick 1907, 511 – 516) – is with Kant's theory of 'Neutral' or 'Moral' freedom of choice, rather than with his theory of 'Good' or 'Rational' freedom.

VI. Non-Moral Freedom

Kant's claim that we necessarily think of ourselves as free, or 'have the idea of' (our own) freedom, is closely bound up with his claim that having the idea of one's own freedom is (or presupposes) being aware of the self-given moral law: 'the moral law is the only condition under which freedom can be known'.[10] But this last claim seems false, taken in one very natural way: it seems that a being that is an intelligent, reasoning purposive agent can acquire the idea of its own freedom (and can have a form of experience which is such that it cannot but form the idea of its own freedom) without having any sort of specifically *moral* sense at all. Its circumstances may be such that it has vivid and repeated experience of being able to do only one of two highly desirable but incompatible things, and of deliberating about which to do – experience of precisely the kind that seems likely give rise to the characteristic absolute sense of being able to choose freely – although it has not acquired any trace of any specifically *moral* habit of thought. If one considers the matter independently of Kant's theory for a moment, it seems obvious that a grasp of moral notions is not necessary for having the idea of oneself as free, even if it is sufficient.[11]

[10] Kant ([1788] 1996, 4 n., Ak.5.4 n). Kant claims that consciousness of the moral law is not only necessary for knowledge of freedom, but also sufficient, for 'if there were no freedom, the moral law would never have been encountered in us' (ibid.). Hence it is the *ratio cognoscendi* of freedom, the means by which we know we are free.

[11] See further Strawson (2021). One cannot plausibly argue, against this, that language, and, hence, membership of a community, is necessary for freedom because necessary for self-consciousness, and that membership of a community of self-conscious agents is in turn sufficient for acquisition of a grasp of moral notions; so that freedom does after all indirectly entail a grasp of moral notions. For, first, a solitary, self-conscious, linguistically endowed,

Nor is a grasp of moral notions necessary for freedom itself. For (a) we don't think that our non-moral choices or actions are in some way less free than our morally weighted choices or actions: if we're free at all, we're as free in the former as we are in the latter. It seems clear that (b) a being with no grasp of moral notions at all can be at least as free in its (uniformly) non-moral actions as we are in our non-moral actions (it has vivid and fully self-conscious experience of facing choices, and so on). But if this is so, then (c) this being can be as free as we are *tout court*. For, generally, (d) if an agent's choice (or action) is free, then, intuitively, this is simply a matter of the choice's (or action's) being *genuinely up to the agent* in such a way that the agent is truly responsible for it (whatever exactly this involves). And whether or not the choice or action is morally weighted is irrelevant to whether or not it is genuinely up to the agent in this way.

It's not irrelevant on Kant's theory (or on parts of it, at least). He explicitly rejects the suggestion that experience of radical freedom to choose is independent of experience of specifically moral obligation.[12] He states that 'the concept of the freedom of the will [of *Willkür*, 'neutral' freedom of choice] does not precede the consciousness of the moral law in us but is deduced [i.e. derived by us] from [our experience of] the determinability of our will [*Willkür*] by this law as an unconditional command'.[13] His argument for this is rather shaky, however, and it doesn't undercut the present suggestion that an entirely non-moral being could have 'consciousness of the freedom of the will' in essentially just the same way as we do, experiencing things as being entirely up to it in such a way that it is truly responsible for what it does. It's a point that is I think well worth noting, if only *en passant*.

VII. 'Will' and 'Desire'

Before setting out Kant's Commitment theory, it's worth commenting briefly on his use of the words 'will' and 'desire'.

---

intelligent purposive agent could logically possibly be created by divine fiat, or occur by chance; second, there is no logical difficulty in the idea of a community of such agents with no grasp of moral notions.

[12] Independent not only in that one can have the former without the latter in a particular case (even if the converse is not true), but also in that there is no general genetic (ontogenetic) dependence of the former on the latter.

[13] Kant ([1793] 1960, 45 n., Ak.6.49). Cf. Kant ([1788] 1996, 30, Ak.5.30); Kant ([1785] 1996, 95, Ak.4.447); Kant ([1781 – 1787] 1933, B, xxxii – xxxiii).

Will. When Kant says that, as an intelligent, purposive, reasoning agent, one believes that one possesses a 'will, that is, …a faculty distinct from a mere faculty of desire' ([1785] 1966, 105, Ak.4.459) he means, as often as not, that one believes that one is able to 'determine oneself to action' (purely) in accordance with reason, and, in particular, in accordance with the self-imposed moral law. But to say that one believes oneself to possess a will distinct from desire may also be to say something simpler and more plausible than this. It may be to say simply that one believes oneself conscious of having a will or power of choice – a power of determining oneself to action – that operates independently of any and all of one's particular given motives or reasons for action, moral or non-moral,[14] a power of deciding how to act which is such that one's choices are not necessarily some possibly complex but nevertheless inevitably determinate function of one's particular motives, but are 'up to one' in a way that is incompatible with their being some such determinate function.

Certainly, we do normally and more or less unreflectively believe ourselves to have such a will and such a power, as remarked in Section III.[15] These terms provide one way among others of describing the characteristic form of our experience of ourselves as free agents. One might say, in Kantian phrase, that we believe ourselves to have a will that is autonomous (or a 'law to itself') just in so far as it fits the description of the will just given, and not because of any sense that we have of ourselves specifically as moral beings. (This can be so even if we can't show that such a will could actually exist, and even if we can show that such a will couldn't exist.) The common experience of being autonomous in this sense seems to be quite independent of any experience that might be said to involve recognition of the moral law in Kant's sense.

To phrase things in terms of the will is to give a quasi-impersonal account of this fundamental feature of our experience of agency. In a way this is unnatural. It may be better to say that the central phenomenological fact is rather that one is a law to *oneself* in being *oneself* a thing that is somehow over and above all one's particular motives. After all, the will doesn't feature as a separate mental thing in one's experience, separate from oneself. The basic belief is

[14] Among motives we may include beliefs that actions are right, and motives that have the character of being apprehensions of duty.

[15] Cf. Hume ([1739 – 1740] 2000, §2.3.2.2): 'We feel that our actions are subject to our will on most occasions, and imagine we feel that the will itself is subject to nothing'.

simply that *I* am, as a choosing and acting being, somehow autonomous – whatever my given desires and motives in any particular case.

Desire. I suggest – very briefly – that we do no violence to Kant's essential intent in taking the word 'desire', in his use, to stand quite generally for any motive with respect to the having of which one is 'heteronomously determined', any motive whose nature is such that, in having it, one counts as being determined by something other than oneself, the rational being that (according to Kant) one most essentially is.

With these comments in place, I will now try to state Kant's – or perhaps 'Kant's' – Commitment theory of freedom.

VIII. Kant's Commitment Theory of Freedom

Having first said that freedom is a 'mere idea', Kant goes on to say that this idea of freedom

> [A] holds good only as a *necessary presupposition of reason* in a being that believes itself conscious of a will, that is, of a faculty distinct from a faculty of mere desire ([1788] 1996, 105, Ak.4.459; my emphasis).

That is, for any being *x*,

> if *x* believes itself 'conscious of a will... distinct from... desire'

then

> *x* necessarily believes that it is free.

This corresponds to step (i) in the commitment-theoretic argument set out in Section IV above: a certain form of experience X, i.e. that which is involved in believing that one has a will distinct from desire, gives rise inevitably to a belief in freedom.

Kant also says that

> [B] every being that cannot act otherwise than *under the idea of freedom* [i.e., in the terms of (1), every being for whom the idea of freedom is a necessary presupposition of reason] is *just because of that really free in a practical respect*.[16]

[16] Kant ([1785] 1996, 95, Ak.4.448), second emphasis mine. There's indefinite scope for quotation and counter-quotation, in support of different interpretations; my concern is to pick out one key strand of Kant's thought.

That is, for any being *x*, if

x is such that it necessarily believes that it is free

then

x really is free (in a practical respect).

This corresponds to step (v) in the commitment-theoretic argument set out in Section IV: the conditions G mentioned in (v) are, for Kant, the conditions of being a self-conscious rational agent; as such, they're identical to the conditions H mentioned in (vii).

Notice that Kant doesn't here say that the idea of freedom is a necessary presupposition of reason in a being that *has* a faculty of will distinct from a faculty of desire, but only that it is a necessary presupposition of reason in a being that *believes* itself conscious of such a faculty. It's not as if the actual possession of some mysterious faculty is being required; only belief that one possesses some such faculty is being required. And no explicit reference has been made to the moral law.

In fact, though, Kant does connect the notion of a will distinct from a faculty of desire with the moral law. For he goes on to say that a faculty of will distinct from a faculty of desire is a 'faculty of determining oneself to action…in accordance with laws of reason independently of natural instincts' (Kant [1785] 1996, 105, Ak.4.459). And by laws of reason he means, in effect, the moral law (or the maxims derived from it). Still, I suggested earlier (Section VI) that one can plausibly understand the notion of a will distinct from a faculty of desire in such a way that belief that one has such a will doesn't necessarily involve any sort of specifically moral awareness, or consciousness of the moral law; and we may adopt this suggestion here. This yields a 'naturalized' version of (1) in which the phrase 'will distinct from a mere faculty of desire' is understood to involve no essential reference to the moral law.

What, then, of (1) – (1) naturalized – and (2)? Together they entail that any being that believes itself conscious of a will distinct from a faculty of desire is (from a practical point of view) really free: for (a) believing oneself conscious of having such a will entails (b) believing one is free, and being unable not to believe this, and (b) entails (c) really being free – so far as acting is concerned ('so far as acting is concerned' corresponds to Kant's 'in a practical respect'). But – it may be said – we are when concerned with freedom principally if not wholly concerned with action. And if we are really free with respect to our actions then we are indeed really free; and if not, not. One can't escape the

consequences of asserting that we are free agents by adding 'in a practical respect'. To say that we are really free in a practical respect is to go further than saying merely that we cannot help thinking we're free, when we act. It is to say that we really are free when we act. It is arguable, then, that Kant has proposed a straightforward positive theory of freedom.

IX. The Inescapability of Belief in Freedom

[B] is the crucial step in the argument, and it's worth trying to exhibit the plausibility – from one perspective, at least – of the suggestion that belief in one's freedom is sufficient for actual freedom, given that one is an ordinary (unhypnotized)[17] self-conscious rational agent. Its implausibility from other perspectives will not be considered here.

Consider the following simple case of Lucia, an ordinary intelligent self-conscious purposive agent. Lucia has a general conception of herself as a free agent. She faces a button. The good and bad consequences of pressing it or not pressing it, in the next twenty minutes, are finely balanced and momentously different. Nothing constrains her. It is (it seems) quite clear to her that she is able to choose freely. But then surely she *is ipso facto* able to choose freely? She may be familiar with determinism, and believe that it is true. But, she reflects, although she will, as soon as the twenty minutes are up, be able to say truly that what she did was determined, that in no way changes the fact that she now knows that she is able, now, to choose completely freely. It is (it seems) an inescapable fact, for her, that this is so. Suppose the choice is a very difficult and painful one. What may oppress her most is precisely her (seeming) knowledge that she cannot escape the choice or her absolute freedom with regard to it: it is hers to make and she must make it.

Situations like this are the heart of freedom as (what one might call) an experiential fact. They're the central facts of the phenomenology – the *cognitive* (i.e. essentially non-sensory) phenomenology – of freedom.[18] It is above all the frequent occurrence in our everyday lives of situations essentially like this one

[17] A man might be subject to post-hypnotic command, or something like it, in such a way that he believed himself free to choose either X or Y, but was in fact unable not to choose ('choose') X.

[18] The term 'cognitive phenomenology' was introduced in Strawson (1986, e.g., vi, 26, 47, 82) to cover all aspects of the character of experience that are more than merely sensory in nature, including all *conative* experiential phenomena insofar as they are more than merely sensory in character. The term is usually now used in a narrower way; see, e.g., Bayne and Montague (2011), Strawson (2011, §12).

(if less dramatic) that makes it seem so impossible that we should ever cease to believe that we are free. And the notion of moral responsibility doesn't have to be mentioned at all. Difficult choices, one could say, are the fundamental experiential guarantors of the belief in freedom.[19]

Suppose Luke's situation is identical to Lucia's, except for the fact that he doesn't have any general conception of himself as a free or truly responsible agent (Kant, of course, holds that this is impossible, given that Luke is a normal human being, a self-conscious rational agent). Suppose he has no conception at all of what it is to be a free agent. Well, then, it seems that he cannot *be* free: it seems that an agent can't be held to be free in the performance of, or truly responsible for, an action, in the normal strong sense, if, when acting, it has no sort of conception of itself as so free or responsible. If this is so, then Luke's possession of the idea of freedom is certainly *necessary* for his freedom, whether or not it is also sufficient given his other properties, as Kant's (or 'Kant's') Commitment theory of freedom suggests. And if so, no theory of freedom can be correct unless it is a Subjectivist theory. But the difficulty in this idea remains undiminished.

X. Another Approach

At one point Kant ties the Commitment-theoretic approach in with his metaphysics of the noumenon. But he does so in a way that suggests that he may have conceived the Commitment-theoretic line to be in some way more fundamental than the noumenal/metaphysical line. He asserts that 'the idea of freedom', which is, for me, a necessary presupposition, '*makes me* a member of an intelligible [or noumenal] world' ([1785] 1996, 100, Ak.4.454; my emphasis), and seems strongly attracted by the idea that it is simply having the idea of one's own freedom that really matters for freedom – given, of course, that one is a purposive agent.

Here Kant seems to be reasoning in the following slightly peculiar way: since true freedom is evidently impossible for a merely empirical being, given determinism, and possible only for an unconditioned, 'intelligible' or noumenal being, and since belief in freedom is sufficient for freedom, in a rational being, it must be the case that belief in freedom actually *makes* me a member of the intelligible or noumenal world. He also says that a rational being is 'involuntarily forced to transfer itself to the point of view of a member of the intelligible world…by the idea of freedom' ([1785] 1996, 101, Ak.4.455). It's

[19] This reflection belongs to part (i) of the Commitment theory sketched in Section IV.

forced to do this simply because it believes it is a free agent and can't help believing this. Significantly, it seems, Kant moves in successive paragraphs from the claim that 'a rational being must *regard* itself as belonging…*qua* intelligence…to the intelligible world' to the claim that a human being as 'a rational being…thus *belong[s]* to the intelligible world' ([1785] 1996, 99, Ak.4.452, my emphasis); as if thinking it is so can actually suffice to make it so – from a practical point of view.

It may be unjust to attribute this reasoning to Kant. He later remarks that the concept of the intelligible or noumenal world is 'only a *point of view* that reason finds itself constrained to take up outside appearances *in order to conceive itself as practical*' ([1785] 1996, 104, Ak.4.458). But there's clearly some havering going on. A few paragraphs earlier the intelligible world is presented as something much more solid, metaphysically speaking, something which '*contains the ground of the sensible world*' ([1785] 1996, 100, Ak.4.453). And it's in the following paragraph that Kant says, *sans phrase*, that 'the idea of freedom *makes me* a member of the intelligible world' ([1785] 1996, 100, Ak.4.454, my emphasis).

It may be suggested that Kant's views on freedom had by the time he wrote the *Grundlegung* evolved, in certain respects, beyond the point at which appeal to the noumenal had much part to play in them. On the other hand, he would doubtless have been quite unprepared to tolerate the metaphysical insubstantiality, so to speak, of the purely Commitment-theoretic approach, and would thus have wished to retain the noumenal metaphysics as serving, in some avowedly incomprehensible way, to ground and substantiate our freedom.

Certainly, Kant seems increasingly clear on the point that true or radical moral responsibility requires that

> human beings themselves must make or have made themselves into whatever, in a moral sense, whether good or evil, they are to become. Either condition must be an effect of their free choice; for otherwise they could not be held responsible for it and could therefore be morally neither good nor evil (Kant [1793] 1960, 40, Ak.6.44).

And since he is of course committed to belief in radical moral responsibility, he holds that such self-creation does indeed take place, and writes accordingly of 'rational beings'…character, which they themselves create' (Kant [1788] 1996, 101, Ak.5.98) and of one's 'knowledge of oneself as a person who…is his or her own originator (Kant [1800] 1993, 213, Ak.22.54). Plainly such radical

self-origination cannot take place in a deterministic order of things, and although it seems clear that it can't take place in any other order of things either, we can draw a veil over this fact by consigning its seemingly evident impossibility to the unintelligible 'intelligible realm'.

XI. A Subjectivist Conclusion

The line of thought just sketched leads to the conclusion that we are free and are so – this is the specifically Subjectivist claim – essentially partly because we believe we are. Commitment theorists take a further step. They claim (perhaps with Sartre) that we are free essentially partly because we can't help believing we are, and not merely because we do believe it. Kant also appears to be attracted by this step. It's clearly important to him that we can't help believing we are free – that it's a *necessary* presupposition of reason for us: it gives a kind of objectivity to freedom in the absence of any other kind of firm metaphysical grounding.[20]

Bibliography

BAYNE, T. – MONTAGUE, M. (2011): Cognitive Phenomenology: An Introduction. In: Bayne, T. – Montague, M. (eds.): *Cognitive Phenomenology*. Oxford: Oxford University Press, 1 – 34.

CASTAÑEDA, H.-N. (1999 [1966]): On the Phenomeno-Logic of the I. In: Kapitan, T. (ed.): *The Phenomeno-Logic of the I: Essays on Self-Consciousness*. Bloomington, Indiana: Indiana University Press, 89 – 95.

HUME, D. (2000 [1739 – 1740]): *A Treatise of Human Nature*. Oxford: Oxford University Press.

KANT, I. (1900 – [1747 – 1804]): *Gesammelte Schriften*. Berlin: De Gruyter.

KANT, I. (1933 [1781 – 1787]): *Critique of Pure Reason*. Trans. by N. Kemp Smith. New York: Macmillan.

KANT, I. (1996 [1785]): *Groundwork of the Metaphysics of Morals*. In: *Practical Philosophy*. Trans. by M. Gregor. Cambridge: Cambridge University Press.

KANT, I. (1996 [1788]): *Critique of Practical Reason*. In: *Practical Philosophy*. Trans. by M. Gregor. Cambridge: Cambridge University Press.

KANT, I. (1960 [1793]): *Religion within the Limits of Reason Alone*. Trans. by T. M. Greene – H. H. Hudson. New York: Harper and Row.

KANT, I. (1993 [1800]): *Opus postumum*. Trans. by E. Förster – M. Rosen. Cambridge: Cambridge University Press.

PERRY, J. (1979): The Problem of the Essential Indexical. *Nous*, 13, 3 – 21.

ROSEN, M. (1988 – 1989): Kant's Anti-Determinism. *Proceedings of the Aristotelian Society*, 89, 125 – 141.

SIDGWICK, H. (1930 [1907]): *The Methods of Ethics*, 7th ed. New York: Macmillan.

[20] This paper is a revised and expanded version of Chapter 3 of Strawson (1986).

STRAWSON, G. (2010 [1986]): *Freedom and Belief*, 2nd edition. Oxford: Oxford University Press.
STRAWSON, G. (2011): Cognitive Phenomenology: Real Life. In: Bayne, T. – Montague, M. (eds.): *Cognitive Phenomenology*. Oxford: Oxford University Press, 285 – 325.
STRAWSON, G. (2014): Freedom and the Self: Feeling and Belief. In: Shoemaker, D. – Tognazzini, N. (eds.): *Oxford Studies in Agency and Responsibility*, vol. 2, *Freedom and Resentment at 50*. Oxford: Oxford University Press, 4 – 11.
STRAWSON, G. (2021): The Phenomenology of Free Agency. In: Erhard, C. (ed.): *The Routledge Handbook on the Phenomenology of Agency*. London: Routledge, 352 – 361.
STRAWSON, P. F. (1962): Freedom and Resentment. *Proceedings of the British Academy*, 48, 1 – 25.

Galen Strawson
Department of Philosophy
University of Texas at Austin
2210 Speedway, C3500
Austin, TX 78712
USA
e-mail: gstrawson@mac.com
ORCID ID: https://orcid.org/0000-0003-0899-5906