

# Conscious Reasoning or Quick Intuition?

*Psychological Accounts of Moral Judgment in between Determinism and Free Will*

Barış Kennedy

## Introduction

In this paper, I will argue that the assumption of determinism in strong psychological accounts of moral judgment leads to methodological issues, which can only be overcome by relinquishing the assumption that such reasoning is non-causal. Central to the argument is the claim that while conscious and causal reasoning may not account for every moral judgment, it is in most cases at least as desirable an assumption as a type of determinism found in strong psychological accounts of moral judgment.

I will begin by defining determinism in the psychological accounts of action and moral judgment, mainly in the work of Jonathan Haidt and Daniel M. Wegner. It will be found that the strong determinism of these models relies on the fact that initial conditions which act as causes of the eventual moral judgment cannot be defined. This leads to serious methodological problems with regard to the construction of models of mental processes. I will argue that even experimental data, which these models frequently use, are not compatible with their deterministic assumptions. I will conclude with a brief discussion of the almost symmetrically opposed ethical foundation of Immanuel Kant, arguing that the assumptions regarding both determinism and free will in meta-ethical psychological models is often a result of the assumptions regarding the accessibility of the decision making project by the judging agent.

## 1. Determinism in Psychological Accounts of Moral Reasoning

Some authors have taken the view that free will and the consequent appeal to reason to set the course of individual action is mistaken, naive and without merit. These authors introduce causes not available to conscious

reasoning to account for the formulation of judgments and actions in which conscious reasoning plays no part. Prominent examples can be found in Wegner (2003) and Haidt (2001). The point is made most clearly by Wegner:

‘We should be surprised, after all, if cognitive creatures with our demonstrably fallible self-insight were capable of perceiving the deepest mechanisms of our own minds. The experience of conscious will is a marvellous trick of the mind, one that yields useful intuitions about our authorship—but it is not the foundation for an explanatory system that stands outside the paths of deterministic causation.’ (Wegner, 2003: 68)

According to Wegner, then, there are deep mechanisms to our minds that function without our knowing, while leading us to believe that we, as agents, are in charge of it all. Our ‘demonstrably fallible self-insight’ is the claim to which Wegner’s proposed model of mental causation of action is intended to lend support. While Wegner’s model is broader in scope than moral action, it certainly has implications for morality in claiming that we are not aware of and therefore unable to change the real causes of our actions.

A model that is directly relevant to moral decision making has been developed by Haidt (2001). Haidt’s ‘social intuitionist model of moral judgment’ intends to answer the question ‘what model of moral judgment allows a person to know that something was wrong, without knowing why?’ (Haidt 2001: 2). To this end Haidt concentrates on the role of intuitions. Intuitions precede reasoning and while moral intuition ‘is a kind of cognition, it is not a kind of reasoning’ (Haidt 2001: 2).<sup>1</sup>

In the intuitionist model proposed by Haidt, the position of moral reasoning and intuitions within the time order is of crucial importance. ‘The central claim of the social intuitionist model is that moral judgment is caused by quick moral intuitions, and is followed (when needed) by slow, ex-post facto moral reasoning’ (Haidt, 2001: 5). While the claim that intuition is quicker than reasoning implies that even if both processes were to begin at the same time, reasoning would be concluded later, Haidt’s claim is stronger: reasoning ex-post facto applies to judgments already made. Reasoning comes after intuition and judgment in the time ordering of Haidt’s model.

This is similar to the position of ‘thought’ in the model proposed by Wegner. In Wegner’s model of ‘experience of conscious will’, thought occupies the position between unconscious causes of action and the unconscious cause of thought that precedes it, and action which comes after thought. The difference between Haidt and Wegner is that for Haidt, ‘reasoning’ begins after judgment has taken place, while for Wegner ‘thought’ precedes the action. In both accounts, however, the real cause of judgment or action precedes conscious thought or reasoning, with neither thought nor reason being a necessary condition for judgment to form.

The similarity between Haidt’s and Wegner’s account does not end at the secondary position of conscious thought within the time ordering. The actual cause which precedes reasoning is unavailable to the subject, at least prior to judgment or action has come to pass. Wegner says as much by labelling these causes as unconscious and stating: ‘The actual causal paths are not present in the person’s consciousness’ (Wegner, 2003: 66). For Haidt, on the other hand, recall that the initial problem appears when people form moral judgments without knowing the reason why they have formed them, that is without being able to express the intuitions which led them to form the judgment. The quick intuitive decisions can sometimes be scrutinized by reason after the judgment has been formed.

The late position of thought/reasoning in the time ordering and the fact that the judgment process is opaque to the agent means that the models of Haidt and Wegner are deterministic in a particular way. Given the initial intuition or unconscious cause of action, the agent is unable to change the action or judgment. The initial conditions set off a causal process that

the agent can do nothing to influence. There might be some conscious thought processes which appear to the agent as having causal power over the judgment or action, but this is either an illusion or an after-the-fact consideration. Hence the determinism of the Haidt and Wegner models has two components: 1) certain mental causes lead to certain judgments or actions and not to others; 2) thought or reasoning has no power to change actions, in fact they are themselves determined by mental causes.<sup>2</sup>

In the following discussion, more attention will be paid to Haidt’s social intuitionist model than to Wegner’s model of illusion of experience of free will. It is important for our purposes to note at this point that Haidt’s account presents a deterministic model, at least in terms of the initial formulation of judgment, which accepts the type of determinism Wegner develops in his arguments against the existence of free will. It will be shown that this two-pronged determinism is incompatible with other requirements of a successful model of mental causation of moral judgment.

Now that we have discussed an important similarity between Wegner’s and Haidt’s accounts, it is important to consider an important difference as well. According to Haidt, it is possible for reasoning to be causal in the narrow sense that it can produce an intuition which is contrary to the original intuition, or produce a judgment. In such cases, reasoning leads to a dual attitude, ‘in which the reasoned judgment may be expressed verbally, yet the intuitive judgment continues to exist under the surface.’ This is the ‘reasoned judgment link’ of Haidt’s model, in which reasoning is truly causal. However, the reasoned judgment link is ‘rare’ and occurs in cases when the ‘initial intuition is weak’ and ‘processing capacity is high’ (Haidt, 2001: 7). Insofar as the ‘reasoned judgment link’ is activated not voluntarily but due to accidental circumstances, I take it to be another component of the ‘complete’ deterministic model.<sup>3</sup>

## 2. The Requirement of Completeness

Any psychological model accounting for judgment such as Haidt’s needs to be able to account for the mechanism by which an individual agent arrives from given initial conditions to a particular outcome and it must

also be able to generalise this mechanism to other agents. Without the first condition being satisfied there is no model, and without the second condition being satisfied the model is not a psychological account of moral judgment as such, but is possibly the study of peculiar behaviour of one individual.

The model can include considerations other than the initial mental cause. We have already seen that under certain circumstances, the slightly different 'reasoned judgment link' becomes active in Haidt's model. It has also been argued that the models of moral judgment proposed by Haidt and Wegner require determinism to be true in the sense that initial mental causes should be exclusive to particular judgments or actions and that reasoning cannot influence the outcome of this causal link, of which it is itself an effect.

In this intuitionist model, given the original intuitions of the individual, the relevance of the moral context to the individual and the background conditions, it must be possible to hypothesise what judgment the individual will arrive at, and whether reasoning will occur after the judgment has been formed (and if so, through which links). But there must also be a general, schematic model, the 'intuitionist model of moral reasoning', which accounts for at least a substantial number of cases of moral reasoning, involving different agents.

Now, it seems that the intuitionist model of moral reasoning accepts that different judgments may be made by different agents under similar conditions. This is not in itself an implausible view. More importantly, however, this finding is already implicated in experimental data Haidt relies on in support of his model.<sup>4</sup>

In accounting for the different judgments of different agents under similar conditions, Haidt's model needs to postulate a difference in terms of the initial intuitions of individuals, or a difference in terms of the influence of links that operate on initially identical intuitions at a later stage. The problem is that since the difference in intuitions tends to determine which links are followed (recall that the reasoned judgment link is activated when the initial intuition is weak and reasoning capacity is high), the actual structure of the model does not seem to make any difference in

terms of explaining why different agents may act differently under similar conditions. Even if different pathways are active in different individuals, this is a consequence of their differing intuitions.

The claim that intuitions are not accessible to the agent at least until a decision has been reached has already been introduced above. This poses a major problem for the claim that different intuitions explain why different agents act in different ways under the same circumstances. The problem is one of measurement but also one of terminological vagueness. Difficulties in defining intuition contribute to the problem of how to experimentally define and differentiate between different intuitions and their effects on different pathways and final judgments. For these reasons, it is unclear what causes the decision making process, which constitutes a methodological problem.

When explanations are given by referring exclusively to intuitions, and any difference between intuitions is supposed to follow from differences between judgments as per the exclusivity of causes, the account presupposes its conclusion as well as the limits it puts on reasoning. Similarly, Wegner's account assumes its own conclusion, which precludes conscious will as a cause. Once various unclear initial conditions of the model are put in the position of not only contributing to the outcome, but also determining the relevant mechanism, reasoning cannot be a cause. It amounts to saying that reasoning has no causal powers, because it is actually at best an intermediary in a mechanism, which is both shaped and activated by something else. We know only that this 'something else' exclusively corresponds to the effect at hand and that and that further inquiry into its nature is futile.<sup>5</sup> Notice that exclusivity and opacity are what render these accounts deterministic in the first place.

A related problem is that just as easily as the psychological model thus comes to reject reasoning as a cause, it can easily account for all kinds of experimental findings regarding moral judgment and action. So we may find that the difference between individuals who find the idea of eating their pet dog repulsive and those who do not (in the Haidt, Koller and Dias study quoted above in footnote 2) can be explained by referring exclusively to intuitions. But even the nature of this difference – whether it is a matter of logical opposition or based on external circumstances – remains a matter of speculation.

I have argued in this section that the requirement that models of moral judgment be complete for all possible judgments across agents does not sit well with the opaque and exclusive determinism of causes in Haidt's psychological model. The danger is that it becomes impossible to falsify the model through counterexamples, as each counterexample of mental causation can be made to fit in the model. However, without completeness within and across agents, the model can hardly claim to account for moral decision making, including those cases in which the agent does not know the causes of her moral judgment.

Below I suggest that the one potential counterexample to the model, a different action of the same agent under the same circumstances, would pose problems to the psychological model in question.

### 3. Inferring Mental Causes and Problems with Completeness

Above I have claimed that psychological models of moral judgment such as Haidt's make the assumption that different agents can act differently under the same circumstances. This is already given in the use of experimental data and, furthermore, accepted in explaining different behaviour through different intuitions. There is nothing that is *prima facie* wrong with the view that different agents can and often do have different intuitions in the same circumstances. However, the model cannot allow for the different action of the same individual under the same circumstances without proliferating its explanatory mechanisms. In other words, there is no room for the assumption that the same individual may act differently under the same circumstances in a given form of the model under discussion.

For an agent to act differently under the same circumstances she must either have different intuitions, or the causal path followed must be different, i.e. through different links to the final judgment.<sup>6</sup> The mechanism remains essentially the same and applies to all moral judgments; the only difference being that multiple paths could follow the same intuition, leading to different judgments.

In this case, any deviations from the expected result, the original judgment of the agent, could be explained by reference to an inter-

vening factor. For example, increased capacity for processing might activate the 'reasoned judgment link' although the initial intuitions are the same or very similar. But to explain the difference in outcome, this interfering factor must be identified, and for purposes of completeness, included in the model. This can only be done provided that the intervening factor is of the same (logical or empirical) significance as the extant parts of the model, the function of which it will replace for the purpose of explaining this case of different judgment. Again the 'reasoned judgment link' provides a good example, since it is genuinely causal. The assumption that the intervening factor is unknown or improbable concedes that the model is not providing a complete description of moral judgment.

In this light, every different judgment formed under identical circumstances requires that an additional causal path be postulated and included in the model. The alternative is to explain why intuitions might differ under the same circumstances.<sup>7</sup>

So far we have argued that the model cannot account for the different judgments of the same individual under the same circumstances, in the sense that the given model needs to alter in structure in order to account for the difference. This is another case of piling on the epicycles. Whenever an explanation is lacking, proponents of such models of behaviour can simply assume that the models are incomplete and add another causal mechanism. Note that the problem could easily be resolved by a more causally potent description of reasoning.

It might be objected that since it is never the case that an agent would have acted differently under the same conditions, psychological accounts such as Haidt's cannot be criticized for their difficulty in accounting for such states of affairs. However, it is another indicator that the model presupposes a form of determinism which denies reasoning causal powers and which assumes a rigid and exclusive correspondence of certain mental causes with certain outcomes. This rejection of free will, or the causal power of reasoning, seems to contribute to the methodological problems of the model, mainly that no plausible counterexamples can be produced even hypothetically.

A possible response to this line of argument is to suggest that the idea of accounting for the different actions of an agent under similar conditions imposes very strict limits and presupposes its own conclusion. Observation and experiment might reveal that there is a statistical correlation between the newly established causal path and its output. By processes of conditioning the interferences with the mechanism may be identified and possibly incorporated into the mechanism. In fact, an evolutionary account, for example, would expect for this to happen and for its mechanism to maintain a statistically significant trend. Hence, without violating the model's legitimacy, such interferences may be applied to cases in which the same agent may have acted differently under the same conditions .

As far as intuitionist models relying on experimentation are concerned, one could reply that such an objection only holds if the interference is as statistically significant as the other parts of the model.<sup>8</sup> In other words, the interfering factor must be just as observable or explainable as admitted parts of the model. To concede this would be to admit that there are certain interfering factors that render, when present, the already accepted mechanism unnecessary. Hence the theory must be improved and either made to account for the genuine cause or must be abandoned in favour of another theory which can do without spurious causes. The model in question simply fails as a causal explanation, if it can always account for the difference of the actions in individual agents under the same conditions. This implies that there are more factors at work in moral judgment than it has taken into account.

The point is that in models such as Haidt's, in which the ethical content of initial causes is unavailable, the model has links which account for different outcomes in terms of the outcomes themselves. It does so by connecting it to a specific process internal to its mechanism. The decision-making process remains unobserved and opaque. At the same time, the number of different mechanistic processes contributing to decision-making may be equal to the number of judgments in need of explanation. Hence, the model will keep repeating itself for every new action of an agent under the same circumstances.

It might be claimed against my objections above that not every moral judgment is made through reasoning and this is an experimentally identi-

fiable finding. One might claim that models such as Haidt's are successful because they propose a mechanism that can explain that and how moral judgment is possible in the absence of reasoning.

Even if we were to accept this limited aim of the psychological model it solves very little. The question of how to account for different behaviour of an agent under the same circumstances now becomes far more pressing, as it seems to be admitted that there is a reasoned outcome and an intuitive outcome. The model has to argue for the cases for which it is relevant.

This can only be done, however, if the question of differing actions of the same agent under the same circumstances collapses into the question of differing actions of different agents under the same circumstances. The empirical data which is cited in support of the model's explanatory power actually consists of cases in which different agents have acted differently under (approximately) identical situations. The desired argument is only possible if, as a result of the different individuals acting differently under (approximately) stable conditions, the model has sufficient information to posit different causal paths or links within the model's structure, enabling it to explain different outcomes. But since the mechanism is supposed to operate at the individual level, the model constructed over the different actions of many individuals is applied to each individual.<sup>9</sup> Agents are then left with the possibility of variant action, which they have not been observed to perform, and which they may never perform.

In response, one might claim that such a criticism may be levelled at any conclusion to be drawn from experimentation and is simply vacuous if one wants to rely on such information for practical purposes. The objection is valid in the following sense: it does not follow that just because someone does not smoke and does not get lung cancer, experimental and statistical data as well as the physiological mechanism which causally relates smoking to lung cancer do not hold for this person. Had the person been a smoker, her risk of lung cancer would have been increased. Similarly, one can claim that people are liable to act more aggressively in an environment with loud noise. However, the issue at stake being whether the model we have developed based on group findings can be applied at the individual level, we must also account for the exceptions, unless it can be unequivocally shown that certain initial conditions always lead to the

same outcome in a representative sample. Continuing the medical analogy, let us say that the exception is not developing lung cancer despite years of heavy smoking. Now it is a case of coming up with an intrinsic mechanistic explanation, as part of the model, for people who have been heavy smokers all their lives and who have not developed lung cancer. Perhaps no such explanation is forthcoming: lung cancer is simply prevented or pre-empted by other circumstantial factors. Or perhaps a genetic factor provides our intrinsic mechanistic explanation. However: either a person has the gene (or any other identifiable factor), which explains why she did not develop lung cancer despite years of heavy smoking, or she does not, in which case we would expect her to have developed lung cancer. Either way, the explanation lies with the individual's constitution. Given our best explanation in the terms outlined: could the individual under the same circumstances (of her life) have (or not have) developed lung cancer? No, we would not expect the outcome to change. Could another person have (or not have) developed lung cancer, given the same life circumstances? Given our genetic explanation, a change in outcome is indeed possible.

This example is intended to show that individual differences cannot be precisely accounted for by applying a model generalised for a group at the individual level. To the question of why a particular individual did not get a disease despite being at risk of contracting that disease, saying some people do not get the disease because of factor x is not an adequate explanation, unless we are also told that this particular individual exhibits factor x. It is at this point that psychological theories of moral judgment or action with opaque and exclusive causes fail; even if we accept that in general there are such unconscious causes of judgment which operate through such and such processes, there is no way in which it can be ascertained that a given judgment on the part of an individual was indeed caused by such a process.<sup>10</sup> These processes include an agent's reflection on her judgment, since the causal process that led to said judgment remains opaque to her. It seems that we are reduced to speculation. A theory that ascribes causal powers to reasoning presents the less cumbersome and the more intuitive alternative. At the very least one can accept without much ado that conscious reasoning may be different in its components (e.g. premises) in different individuals, resulting in differences in judgment.

#### 4. The Failure of the Deterministic Psychological Model

This is not the place to argue the relative merits of a concept of reasoning that presupposes the existence of free will and determined intuitions; it is sufficient to observe that Haidt's model does not succeed in producing a better explanation of moral judgment than the rationalist view it criticises.

Haidt's model may be able to explain peculiar cases of moral uncertainty, cases in which the agent is certain of the moral judgment, but cannot account for it or give reasons. In Haidt's words: 'In the social intuitionist model it becomes plausible to say I don't know, I can't explain it, I just know it's wrong.' (Haidt, 2001: 2) Anyone who professes this opinion has, according to Haidt, formed a judgment through intuitions, which cannot be corrected or argued for, and is stuck with avowing the normative value placed on that which is wrong.

This theory of unreasonable moral judgments, however, renders every moral judgment uncertain, possibly over the protestations of the judging agent herself. We have seen that the account is very troubled methodologically. Its deterministic assumption of opaque, exclusive causes leading to judgment with no possible interference by reasoning leaves nothing unexplained and is impossible to dispute, while being very difficult to empirically verify.

In conclusion the time-ordered deterministic intuitionist model of Haidt fails to deliver an acceptable account of how a range of moral judgments are produced in individual agents. This remark has certain corollaries. Firstly, determinism is latent in Haidt, but as this essay attempts to display, this is precisely the problem. In the light of opaque and exclusive determinism of causes, the system of modelling used simply cannot admit of a single type of cause, such as intuitions, without running into methodological troubles. The limited role of reasoning in Haidt, which make it depend on intuitions even when truly causal, presents reason only as an ad hoc consideration, which renders the theory even less falsifiable. My proposal is that the assumption of determinism be made explicit and a correction be made to the overdetermination of both causal and time-ordered positioning of intuitions and reasoning, so that reasoning can precede some intuitions. The model may be made more specific for certain types

of moral reasoning and judgement and allow for the agent's awareness that decision-making is a mental process even in those cases. As I hope to have shown, it is a tall order to accept at face value a statement 'I don't know why, I just know it is wrong' if 'I don't know' why here means 'I cannot elaborate any further on the causes'.<sup>11</sup>

I want to suggest that the fundamental problem of the psychological account of Haidt is that judgments and intuitions are completely interchangeable; they do not differ in content. Not only is there no valid or invalid judgment, true or false intuition; there is also no distinction between moral judgments in terms of type (as one finds in Kohlberg, whom Haidt quotes extensively). It is suggested that reasoning may supplement such categories, but of course, it hardly plays a part in forming judgments. Since without knowledge of its content we have no way of establishing that intuition  $x$  leads to judgment  $y$  (as if one could be logically derived from the other) we have to accept the very troublesome requirement of exclusive causes of each particular judgment at face level. This condition then necessitates determinism in the sense that (1) given the particular initial conditions there can be no different outcome and that (2) reasoning cannot be causal.

Interestingly, the connection between aspects of moral decision-making and the question of free will can run in the completely opposite direction. A meta-ethical theory which does not separate between the grounding of moral judgement and the form of morality itself runs the opposite difficulty of having to insist on an absolute conception of free will. Immanuel Kant's practical philosophy exemplifies this later point.

## 5. Kant and the Moral Law

I want to briefly examine the Kantian position in order to claim that theories on judgment forming have to confront the problem of free will and determinism. It seems with regard to this problem, Kant's position does not follow from a priori statements about freedom, but by the requirements of his meta-ethical model regarding the structure of judgment making. Of particular importance is the agent's awareness of the act of judging and to what extent this is an essential part of judging. We have

seen that the psychological models of Haidt and Wegner render this very process unconscious or otherwise unavailable to the mind and converge on determinism. The opposite choice is to be found in Kant, for whom our knowledge or freedom is guaranteed by apperception. This idea of the possibility of ethics being implied by ethics itself constitutes an identification of ethics with meta-ethics.

In the preface to the Critique of Practical Reason Kant famously declares:

'Lest anyone should imagine that he finds an inconsistency here when I call freedom the condition of the moral law and hereafter maintain in the treatise itself that the moral law is the condition under which we can first become conscious of freedom, I will merely remark that freedom is the ratio essendi of the moral law, while the moral law is the ratio cognoscendi of freedom.' (Kant, 1788/1996: 140)

This, of course, presents a stark contrast with the investigation of the concept of freedom as an antinomy in the Critique of Pure Reason. Freedom, having been denied theoretical value in laying down the foundations of knowledge for pure reason, now makes a comeback, despite Kant's insistence in the first Critique that: '(...) and if we could exhaustively investigate all the appearances of men's wills, there would not be found a single human action which we could not predict with certainty, and recognise as proceeding necessarily from its antecedent conditions'. (A549-550/B577-8, translation BK)

In his book 'Kant's Theory of Freedom', Henry Allison points out that in this passage 'Kant is advocating not only a strict determinism at the empirical level but also a psychological determinism' (Allison, 1990: 31; emphasis in the original). Why did then Kant change the outlook he had established in his first Critique when it came to morality, leading many commentators to draw a sharp distinction between his theoretical and practical philosophy, often at the detriment of the latter?

As indicated above, I claim that this was the result of Kant's identification of meta-ethics with ethics, which arose from his view of reason. While it is not the aim of this paper to offer a comprehensive interpretation of

the Kantian practical philosophy, it must be pointed out that reason in this context occupies a pivotal role, as the starting point of every ‘model’ which can explain behaviour. This is the role played by apperception, which means being aware of the operation of a faculty of reason, not retrospectively but as a necessary consequence of such an operation as judging. Any judgement in itself might appear necessary, and theoretically it is; for it is made either in accordance with the categorical imperative or with nature. However, apperception allows for an explanation of, or at least reasoning on, how a judgement comes about. It is appropriate to quote at length from the Critique of Practical Reason here to show the significance of this point of view:

‘Whether he would or not perhaps he will not venture to say; but that it would be possible for him he would certainly admit without hesitation. He judges, therefore, that he can do something because he knows that he ought, and he recognised therein his freedom, which, without the moral law, would have remained unknown to him.’ (Kant, 1788/1996: 163-164)

Notice that the consideration of doing what is right establishes the moral law, which in turn establishes freedom, which, as we saw above is the *ratio essendi* of the moral law. Fine points of transcendental deduction aside, what lies at the root of this reasoning is the truism, ‘I may or may not do something I know to be right’.

In Haidt we saw that any judgment can be explained in terms of original intuitions and accepted to be a moral judgment of the agent, even if the agent has no insight into the formation of that judgment. There is no distinction between types of judgment. In Kant we see that judgement has the potential to be moral. Here lies the difference between the assumption of determinism in one case and the assumption of freedom in the other. Kant’s meta-ethics is concerned with identifying the necessary grounds for the possibility of correct judgement according to a certain conception of the ethical. By way of contrast, Haidt’s indifference to the ethical content of judgment allows him to account for all judgments as a purely cognitive type, determined by the uniform nature of cognition.

Ethical considerations such as the categorical imperative inform the Kantian view of what judgments may be considered under the moral category and therefore as free, in the Kantian sense. While there is no reason to uncritically accept this point of view, it is still able to answer methodological questions like those we have wrestled with above, such as the ability to account for the different action of individuals in the same circumstances. Assuming free will seems to make it easier for the Kantian position to allow for every instance of moral judgment admits of the same mechanism or process which is uniform across individuals.

It would be misleading to suggest that Kant’s abandonment of his theoretical view of freedom in favour of a new one for his practical philosophy was unproblematic. As mentioned, most critics were rather unimpressed with Kant’s ‘deduction of freedom’. His views on the empirical as necessarily determined and the knowledge of freedom for moral purposes led him to what is called the ‘two-world view’, where the agent is considered to be empirically causally determined but noumenally free. This is a very controversial proliferation of ontology; furthermore it is problematic in its transcendental idealist assumptions.

## Conclusion

The comparison between the assumed free will in the Kantian position and the structurally problematic determinism of the models of Haidt and Wegner exposes a dilemma for strong psychological accounts of moral judgment. The problem is whether to admit of the relationship between the content of moral intuitions and judgment as influenced by reasoning or not. If we admit of such reasoned content as ‘incest is wrong whatever the outcome; it is wrong even if no one got hurt’, the focus turns towards the structural analysis of reasoning and the question whether the judgment is sound, in order to show that it ought to be preferred to ‘any action is permissible so long as no one gets hurt’. We will be engaged in the practice of ethics and a typology of judgments insofar as there are clearly measurable individual differences between groups of such judges, in meta-ethics such as the one practiced by Kohlberg. No mechanism will be available to explain situations in which the agent insists on a judgment despite all



argument to the contrary. But if we decide to account for these cases, we have to turn to an account of moral judgment which is indifferent to the content of judgments. This in turn relies on an unwelcome type of strong determinism with severe methodological handicaps.

*Barış Kennedy was born in Ankara, Turkey in 1985. He graduated from Hacettepe University with a BA in philosophy in 2009 and took his MA in philosophy at Erasmus Universiteit Rotterdam in 2012. His current research interests include history of early Enlightenment thought and philosophy of mind.*

*'Conscious Reasoning or Quick Intuition?' was written for the master course 'Experimenting Ethics Away' (of prof. dr. Maureen Sie) taught by dr. Leon de Bruin.*

## Notes

1. Intuitions form the mainstay of Haidt's social intuitionist model and have their own 'link' or 'process': the intuitive judgment link. Other types of moral cogitation fall into five other types of link, but it is important to note that moral judgment always starts with the initial intuitive judgment link (Haidt, 2001: 7).
2. The first point on one to one correspondence between mental causes and their outcomes may not seem obvious. The point is that since these accounts begin with observations of action or communicated moral judgment, and since they pose a set of mental phenomena as the causes of observed behaviour, they have no choice but to assign different causes to different effects exclusively.
3. The main function of reasoning in Haidt's model is not as a cause of moral judgments, but as a means of justifying decisions to one's self and persuading others of the value of the moral decision. I do not treat of the 'social' aspect of Haidt's model.
4. For example, regarding the Haidt, Koller and Dias (1993) study cited by Haidt (2001) on page 5, Haidt says 'The stories were carefully constructed so that no plausible harm could be found, and most participants directly stated that nobody was hurt by the actions in question. Yet participants still *usually* said the actions were wrong, and universally wrong' [emphases mine]. Though we are not given the proportion of participants who may or may not have acted in these ways, presumably some participants judged it best not to say that the actions did not hurt anyone and others may have even judged that the actions were not universally, or otherwise, wrong.

5. Admittedly in Haidt, the intuition can be investigated by retrospective reasoning, but it cannot be changed, or even understood in its role within the causal chain. It can only be voiced. In Haidt there is an underlying assumption that the intuition, propositionally expressed, is the same as the concluding proposition of a line of reasoning. Only in terms of functioning do intuitions and reasoning differ. (see Haidt, 2001, Table 1).

6. Since the model seems to assume every different outcome to have a different cause, the agent must have different intuitions under the same circumstances. This is another consequence of the vagueness of the initial cause in Haidt's model. But I will assume for argument's sake that an intuition or sufficiently similar intuitions may still follow different causal paths to different judgments.

7. This presumably calls for an analysis of whether intuitions have stable qualities in individuals over time. While the possibility of such an analysis is not to be ruled out, for the theoretical burden intuition bears in these psychological models, its absence is certainly a negative point.

8. Since claiming the effect of unknown or improbable interfering factors would be just another way of claiming that the theory does not work for unknown reasons.

9. The use of the first person plural in talking about such models serves this confusion. Sentences such as 'We frequently form negative opinions due to negative environmental factors without being aware of this' abound in the literature. 'We' as in me, you, him and her or 'we' as in the experimental sample?

10. There is the further point that in our smoking causes lung cancer example the preventative genetic factor is introduced as a causal factor without any strong assumption of determinism. In this case it is simply a sufficient reason. Intuitionist determinist models, on the other hand, need to point at the determining power of the intuition and the resulting causal chain, without clear empirical evidence for such a mechanism. Again we come to the issue of inferring a mechanism from observed behaviour, where the proof of the mechanism does not seem forthcoming. It is telling that the medical establishment only accepted the connection between smoking and lung cancer after the mechanistic evidence of how tobacco smoke damages the lungs was given, despite there being good statistical evidence beforehand (Machamer et al, 2000). This is a more general point about methodology, compared to the problem of applicability at the individual level given above.

11. Rationalisation, which Haidt sees as the main function of reasoning, need not always produce logical arguments from admissible premises. The participant who reacts to the vignette about incest in which neither of the siblings is harmed in any way (Haidt, 2001:2) can still successfully rationalise his objection with the following two premises: 1) Incest is taboo; 2) What is taboo must not be done.

## References

- Allison, H.E. (1990) *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.
- Haidt, J. (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment, In: *Psychological Review*, 108, 814-834.
- Kant, I (1788/1996) *Critique of Practical Reason*, translation M.J. Gregor & A. Wood. Cambridge: Cambridge University Press.
- Kant, I. (1998) *Critique of Pure Reason*. Cambridge: Cambridge University Press.
- Kohlberg, L. (1973) The Claim to Moral Adequacy of a Highest Stage of Moral Judgment, In: *The Journal of Philosophy*, 70(18), 630-646.
- Machamer, P., Darden, L. and Craver, C. (2000) Thinking About Mechanisms, In: *Philosophy of Science*, 67, 1-25.
- Wegner, D.M. (2003), The Mind's Best Trick. How We Experience Conscious Will, In: *Trends in Cognitive Sciences*, 7(2), 65-69.

