

## ABELARD'S INTENTIONALIST ETHICS\*

### Introduction

**A**BELARD'S ethical theory, as presented in his *Ethics*,<sup>1</sup> is a version of what I'll call 'intentionalism': the view that the agent's intention determines the moral worth of an action. Now even in Abelard's day, the common understanding of morality<sup>2</sup> seemed to endorse the following principle:

(P) An agent should intend to  $\varphi$  only if bringing about  $\varphi$  would be good  
But Abelard replaces (P) with its obverse, a principle he identifies as the rational core imbedded in traditional Christian moral teaching:

(P\*) An agent should bring about  $\varphi$  only if intending to  $\varphi$  would be good  
Abelard's arguments against (P) and in support of (P\*) are remarkably similar to those given by the most famous exponent of intentionalism: Kant. For Kant's ethical theory, especially as he presents in the first section of the *Groundwork of the Metaphysics of Morals*,<sup>3</sup> identifies (P\*) as the philosophical conception corresponding to the "common rational knowledge of

<sup>1</sup> Abelard's three ethical works are (i) his *Commentary on "Romans"*, cited from the critical edition of Eligius M. Buytaert, ed. *Commentaria in Epistolam Pauli ad Romanos*, in *Petri Abaelardi opera theologica, Corpus christianorum: continuatio mediaevalis* tom. XI, Turnholt: Brepols 1969; (ii) his *Dialogue Among a Christian, a Jew, and a Philosopher*, cited from Rudolf Thomas, ed. *Petrus Abaelardus: Dialogus inter Philosophum, Iudaeum, et Christianum* (Textkritische Edition), Stuttgart-Bad Canstatt 1970; and (iii) his *Ethics*, cited from David Luscombe, *Peter Abelard's "Ethics"*, a critical edition with English translation, notes, and introduction, in the Oxford Medieval Text Series, Clarendon Press, Oxford 1971. Translations from these and other works by Abelard are mine.

<sup>2</sup> Despite widespread agreement on particular normative principles in Abelard's time, derived from the Bible, and despite the then-current view that education dealt with the formation of the moral character of the student, there was no systematic treatise on ethics. The Church Fathers wrote about theological virtues and their role with regard to grace and salvation, but not about systematic ethics; Aristotle's *Nicomachean Ethics* would not be translated for another century; the epistles of Seneca, who in the twelfth century was mistakenly believed to be a correspondent of St. Paul (and so a crypto-christian), offered a full range of moral advice with only the sketchiest of hints about their underlying theoretical structure. Abelard's accomplishment is all the more impressive in this light.

<sup>3</sup> I'll refer to the Akademie version of Kant's *Grundlegung zur Metaphysik der Sitten*, citing the page numbers thereof, but most quotations will be drawn from Lewis White Beck's translation (Bobbs-Merrill 1959).

morals.” Abelard and Kant locate moral worth in features of the way the agent conceptualizes her performances, and each thinks that goodness is characterizable in terms of the form such conceptualization takes. Both are deeply indebted to Stoic ethics, familiar to each largely through Seneca’s letters, and they share a common project: ‘christianizing’ Stoic metaethics so that the classical equation of virtue with happiness is revised to leave room for God and the Afterlife.

I’ll proceed as follows. In §1, Abelard’s arguments against (P) will be canvassed. In §2, I’ll look at his arguments in favor of (P\*) and a ‘mediæval categorical imperative’. In §3, a comparison with Kant’s intentionalism should make the virtues of Abelard’s theory apparent. In §4, I’ll discuss Abelard’s failed attempt to baptize Stoicism. Finally, by way of conclusion, I’ll offer a suggestion about why intentionalism, in any of its versions, is an ethical theory to take seriously.

### 1. The Source of Moral Worth

Abelard’s method in the *Ethics* is straightforward. He isolates four factors involved in the performance of a deed that are candidates for the bearer of moral worth (*i. e.* that which determines the moral quality of the deed). These factors are the desires of the agent, the agent’s character, the deed performed, and the agent’s intentions.<sup>4</sup> He eliminates each of the first three candidates, typically by a lively use of counterexamples, and by so doing presents the negative case against (P), leaving the field clear for his positive account of (P\*).\*<sup>5</sup>

<sup>4</sup> The last factor Abelard analyzes into two components: the *consensus* of the agent to a desire which is already present, and the *intentio* of the agent to act. Abelard isn’t very careful about the distinction between consent and intention. It seems as though an intention is a here-and-now desire to  $\varphi$ . Consent to a desire seems to produce an intention, which we may think of as a here-and-now desire to  $\varphi$  that has motivational force. But it is an odd sort of ‘desire,’ since it is intellectual. It is not clear to me that Abelard has a theory to back up his admittedly intuitive use of ‘consent’ and ‘intention.’ It may be that the sort of ‘desire’ that is involved in intending is rather like Kant’s notion of ‘respect’: neither is a desire on a level with, say, the desire to drink when thirsty; each is used as an ‘intellectual’ version of desire, having motivational force; each is free from the strict determination which characterizes bodily desires, being solely in the power of the will (or the intellect and will together). This is perhaps a place where Kant can provide a theoretical justification for Abelard’s practice. I’ll simply follow Abelard in his intuitive use of the terms.

<sup>5</sup> As I sketch Abelard’s arguments I will not, in general, respect his order of presentation; they can be summarized in a compact form which should make their forcefulness apparent. References to Abelard’s arguments, unless otherwise noted, are to his *Ethics*.

### 1.1 Moral Worth Does Not Come From Desire

Abelard gives a three-stage argument to eliminate the first alternative, that the desires of the agent determine the moral worth of the action.<sup>6</sup> First, Abelard argues that some deeds pre-theoretically taken to be evil can be performed without any evil desire.<sup>7</sup> He establishes this by an example of self-defense (6.24–29):

Consider some innocent man whose cruel lord is so furious at him that he chases him, brandishing a sword, to kill him; that man flees as far as he is able to avoid his own murder, yet finally he unwillingly kills (his lord) lest he be killed by him. Tell me, whoever you are, that he had an evil desire in this deed!

Killing is evil, yet the innocent man of Abelard's example had no desire to kill his lord, but merely to preserve his life. No evil desire is present. Abelard cautions us against confusing the following desires (8.21–26):

The desire to  $\varphi$ -for-the-sake-of- $\psi$

The desire to  $\varphi$  (*simpliciter*)

The latter desire is not entailed by the former. I may desire that you have the shirt off my back for a king's ransom, but that doesn't entail that I simply desire to give the shirt off my back. In fact, the only inference which can be drawn from the former sort of desire is that the agent desires  $\psi$ . But this is compatible with the agent not desiring  $\varphi$ , or desiring the opposite of  $\varphi$ ; Abelard says that in such cases the agent "endures what he does not wish for something he does desire."<sup>8</sup> His desire is praiseworthy in itself,

<sup>6</sup> Abelard's argument is somewhat obscured for the modern reader since he uses the term '*voluntas*' for desire, often misleadingly translated as 'will' (Luscombe translates it in this way). Abelard clearly does not mean to be talking about a faculty of the soul, or about particular volitions, but about desires in something like the modern sense. In his *Commentary on "Romans"* 207–209 Abelard distinguishes between *velle* in the sense of *voluptas*, *desiderium*, and *delectatio*, and in the sense of *approbare*. The former senses are relevant to the *Ethics*, and so I shall translate *voluntas* as 'desire' throughout.

<sup>7</sup> Abelard actually says that we sin (*peccare*) without evil desire. His choice of terms is poor: he does elaborate a theory of sin, but it would be question-begging to appeal to that theory in the course of argument. Yet it is clear that he only needs to point to a pre-theoretical moral evaluation of a deed, and accordingly I will paraphrase him in this way.

<sup>8</sup> The situation is surely more complicated than Abelard allows. If an agent desires to  $\psi$  and believes that doing  $\varphi$  will bring it about that  $\psi$ , then there is a sense in which the agent can be said to desire to  $\varphi$ . Part of the difficulty comes from Abelard's failure to distinguish between processes (or engaging in processes) from results (or producing

since Abelard holds that there is a moral imperative to preserve one's own life.

It might be objected that situations involving internally complex desires or actions under constraint, such as self-defense, are exceptional; the prohibition against killing is relaxed in such circumstances, and it is illegitimate to conclude something about ordinary cases from these special circumstances. Abelard recognizes this objection (10.28–30), and the second stage of his argument is to argue that there is no merit in simply doing as we desire, regardless of the deed performed. For “what great thing do we do for God if we support nothing against our desires but rather do whatever pleases us?” (12.14–15). Actions are not praiseworthy if they simply satisfy the desires of the agent. A person who gives alms to the poor not because he believes it is good but because he likes giving money to people, and the poor are willing to take it, is not praiseworthy. This point can be generalized. In order to be morally evaluated, a deed has to be performed for the sake of something; that is what makes a deed an action. Actions are evaluated in terms of their ends or purposes. But satisfying desire is not a morally valuable end or purpose—perhaps because our desires are no secure guide to right action, perhaps because it does not embody the appropriate impartiality or impersonality characteristic of moral action. Mere satisfaction of desires will not do, no matter whether the content of the desire be selfish or altruistic.

Abelard argues for an even stronger claim, namely that the moral value of an action is enhanced if the agent performs it without desire, or if it is against the agent's desires. As he says (12.3–6):

If desire is restrained, though not extinguished, by the virtue of temperance, [the desire] remains for a fight and persists in struggling and does not give up even when overcome. Where is the fight if the material for fighting is absent? Where does the great reward come from if what we endure isn't difficult?

Abelard illustrates his position with an allusion to Augustine: if someone passing through someone else's garden should see some fruit on the trees, then if he has no desire for the fruit, he is not praiseworthy for leaving it undisturbed. However, if he “falls into longing” for the fruit, then his passing them *is* praiseworthy—and the stronger the desire, the more praiseworthy passing them by (14.4–13).<sup>9</sup>

Desires, therefore, cannot be the sole determinant of moral worth: ac-

results) as the objects of desire and intention.

<sup>9</sup> This argument is no more convincing than Kant's grocer-example, and for much the same reason: each trades in an unacceptable way on the content of the desires in ques-

tions which would pre-theoretically be judged evil can be engaged in without any evil desire; actions solely on the basis of desires have no moral value at all; actions which would pre-theoretically be judged good would have their value enhanced if engaged in without desire or against one's desires.

### 1.2. Moral Worth Does Not Come From Character

Abelard holds that character traits are simply complex patterns of mental dispositions of desire and feeling (2.21–22). To be irascible, for example, is to be prone to or ready for the emotion of anger. The previous rejection of desires as determining moral worth immediately leads to rejecting character as determining moral worth—since desires themselves lack moral value, so *a fortiori* dispositions-to-desire lack moral worth. There are no facts about the dispositions that could make them different, in the morally relevant way, from desires.<sup>10</sup>

Abelard offers an additional argument against character traits as determining moral worth. It is a fact that good and bad men can have much the same set of character traits; thieves can be courageous, honest men intemperate. But whatever can “occur in both good and evil men is not relevant to morality” (2.13–14). Any characteristic present in good men which is present in evil men cannot be that which makes the good men good since its presence in the evil men would make them good.<sup>11</sup> Character

tion. If the example were rather of a person who has strong desires to aid the poor and homeless, but managed to struggle against and overcome his desire to give them aid, we would be far less inclined to endorse the conclusion that moral value is enhanced by the struggle. Abelard's position suggests a rather unqualified endorsement of asceticism, widespread in the twelfth century, but less palatable to the twentieth century. Note also that the position suggests that one should not eradicate desires but encourage them so as to nearly be overcome by them—for the sake of the struggle! Abelard comes dangerously close to holding this thesis (12.9–13).

<sup>10</sup> This argument should be qualified: it is not logically impossible to hold that desires cannot determine moral worth but that the possession of certain dispositions is morally valuable. Such a view would be rather peculiar, especially given that consent to desire and not desire itself is the relevant question. If character traits were analyzed not as dispositions to desire, but as dispositions to consent, the issue might be very different. Abelard suggests as much in the second book of the *Ethics*, when he turns to considering virtues and vices as acquired dispositions, but unfortunately the rest of the manuscript has been lost (if it ever existed). See *Ethics* 128.18–30.

<sup>11</sup> Or so Abelard seems to think. To make this line of reasoning cogent, we at least have to add the proviso that we consider the set of all characteristics, since it is not unreasonable to think that the property  $y_1$  combined with  $y_2$  could make the possessor of both good, whereas  $y_1$  combined with  $y_3$  could make the possessor of both evil. It might be thought, of course, that  $y_1$  is not what makes the good man good, since it

in itself cannot ground moral worth.<sup>12</sup>

### 1.3 Moral Worth Does Not Come From the Deed

Abelard attacks two ways in which the deed might be taken to ground moral worth. On the one hand, deeds are sometimes evaluated and justified on the basis of their purpose or their point; on the other hand, they are evaluated and justified in terms of their intrinsic nature or the consequences that flow from them.

Abelard's argument against the purpose of a deed is simple: take any deed for any given purpose, and you'll be able to imagine a case in which the deed is performed for that purpose but the agent's intention is evil. He offers two examples. First, Judas and Jesus each performed deeds with the same purpose: to bring it about that Christ be crucified. But Judas's deed was evil, whereas Jesus's was not (28.2–9);<sup>13</sup> more generally, Satan does nothing but what God permits, and so the same deed with the same purpose (*e. g.* causing Job misery) is evil with respect to Satan but good with respect to God (28.18–24). Second, Abelard considers a situation in which the deed and the purpose of the deed is identical for each of two agents, but distinct intentions require us to render distinct moral verdicts (28.11–17; see also *Dialogus ll.* 3267–3272):

Often the same thing is done by different people, [but] done through the justice of one and the iniquity of the other. For example, if two men hang a convict, one out of his zeal for justice and the other from the hatred stemming from an old enmity, although the act of hanging is the same and each does what it is good to do and what justice requires, nevertheless the same thing comes about through the difference in [their] intentions [so that] by one it is done well and by the other badly.

The deed is identical and the purpose identical, but moral worth depends on the intention of the agent(s) involved.

is at most a partial cause of his goodness, but Abelard needs to provide an argument for this claim.

<sup>12</sup> Another kind of argument for this conclusion, which Abelard does not consider, is how to evaluate action which is 'out of character'. It is possible to hold the view that the moral worth of an agent stems from his character, but the moral worth of an action depends in a complex fashion on the agent's character and the particular motives of or reasons for any action.

<sup>13</sup> To insist that Judas did what he did for thirty pieces of silver while Jesus did what he did to redeem mankind is to appeal to their intentions, and so to grant Abelard's point that the deeds themselves only have moral worth only through the intentions they embody.

Nor will it help if we try to relativize evaluative terms to the 'point' of the deed, as some have taken Aristotle to do, so that the assessment of a deed depends on whether it is a good or bad instance of that type of deed. Just as a knife is good or bad *qua* knife if it does well or poorly at the things for which knives are designed, so too we might think that deeds embody evaluative criteria relative to the kind of deed they are.

In *Dialogus ll.* 3254–3260 Abelard argues that this relativization of evaluative terms results in terms that are fundamentally non-moral: the deed specified by the description 'baking a cake' can be performed well or badly, it is true, but this is the case for any deed under any description. Robbing a bank can be done well or poorly, as can murder. Rather, the moral worth of the content of the description is what matters, and it is only derivative upon this that the purpose of the deed has moral worth. Events, however, fall under a variety of descriptions. Which descriptions are relevant to the moral evaluation of the deed? At a minimum, it seems as though the agent's intentional description is an important factor—but that is just to import intentionality, which is Abelard's aim.<sup>14</sup>

To show that the deed and its consequences or effects do not determine moral worth, Abelard begins by criticizing the alternative: the position that the performance or non-performance of deeds is *all* that matters, a 'strict liability' ethical theory. This alternative might be thought especially attractive to traditional Christian teaching, since it proceeds by way of commandments: absolute prohibitions regarding performance and non-performance, such as 'Thou shalt not kill.'

Abelard's first objection to a strict liability theory is that such commandments, construed only with regard to the deed, fail to condemn those who are obviously evil, namely those who have nothing but the worst of intentions yet are never in a position to act on them. (This is a plausible generalization of Abelard's case of the willing perjurer who does not get to perjure himself, 26.8–14). His second objection is that nobody can keep from violating such prohibitions. Abelard offers a version of the story of Oedipus: fraternal twins, male and female, are separated at birth and neither learns of the existence of the other; as adults they meet, fall in love,

<sup>14</sup> Abelard's argument is too weak: while the agent's intentional description of the deed is relevant, it might ultimately be disregarded in evaluating the deed. (The distinction between an event and an action is important to Abelard's argument here.) An agent may think that unplugging the hospital machine is an act of mercy, whereas we may decide that it is more correctly described as a case of murder. Hence Abelard needs to show that it is the intention of the agent which makes all the difference, and the argument isn't strong enough to warrant this conclusion.

are legally married, and have sexual intercourse.<sup>15</sup> Technically this is incest, but Abelard finds no fault in either to blame (26.14–23). If the deed alone determines moral worth, then on a strict liability theory their (justifiable) ignorance is morally irrelevant—which it manifestly is not.<sup>16</sup> Absolute commandments, Abelard concludes, deprive the actor of any status as a moral agent.

Abelard expands his attack on the deed and its consequences with a pair of cases centering around what recently has been called ‘moral luck’: cases in which nonmoral factors enter into or affect the possibility of moral actions. His first case is that of hypocrites and the wealthy. Such individuals are far better motivated (by the love of praise) and situated (by their riches) to perform acts that have wide effects and far-reaching consequences than the ordinary individual. But surely these aren’t morally relevant factors, even if views taking the deed to be the sole determinant of moral worth must count them as such (28.24–26). Abelard’s second case has to do with two men, each with the money and intention to build poorhouses; the first is robbed before he can act, while the second is able to build the poorhouses. To maintain that there is a moral difference between the two men is, Abelard says, to hold that (48.21–28):

...the richer men were, the better they could become. To think this, namely that wealth can contribute anything to true happiness or to the worthiness of the soul, is the height of insanity!

Deed-centered morality loses any counterfactual purchase on what might have been the case, and so cannot separate moral and non-moral factors. To allow for the possible and the might-have-been, something other than the actual deed has to enter into the determination of moral worth. Character, as a pattern of dispositions, might be thought to provide such counterfactual purchases, but Abelard has ruled this out earlier (see §1.2 above). Intentions, however, need not be discharged for the agent to be praised or blamed (14.17–19), and so do provide counterfactual purchase for moral assessment.

Abelard then attacks the very coherence of the notion that deeds could be the determinant of moral worth. He enunciates a dualist principle (22.31–

<sup>15</sup> Ignore the question whether they are in fact married if they are so closely related—nothing in the example turns on it.

<sup>16</sup> Another interpretation of Abelard’s objection is stronger. According to the view that the deed alone matters, it’s plausible to think that agents should refrain from performing any action that might violate or lead to a violation of the prohibition. Then Abelard’s point would be that any human action involves a nonzero probability of violating an absolute prohibition, either in itself (as through the unwitting commission of incest) or in its consequences (for want of a nail the battle was lost).



34):<sup>17</sup>

As if what is exterior and physical could contaminate the soul! The performance of deeds is in no way relevant to the increase of sin. Nothing stains the soul except what pertains to it.

Abelard takes this principle to license a claim that can be defended on its own merits, namely that deeds (and their consequences) are in themselves indifferent:

[*Ethics* 44.30–32]: Deeds, as we have said, are common to good and evil men alike; in themselves they are indifferent, and should only be called ‘good’ or ‘evil’ on account of the agent’s intention.

[*Dialogus* ll. 3158–3163]: I hold a thing to be *indifferent* which is neither good nor evil. . . Actions should only be judged good or evil with regard to their root, the intention [of the agent].

Deeds are not subjects of moral evaluations; only agents are. Abelard derives this claim from Stoicism, which notoriously maintained that deeds in themselves are indifferent (*ἀδιάφορα*).<sup>18</sup>

Abelard argues that the performance or non-performance of the deed doesn’t alter moral worth in any way—and in particular it doesn’t add to or detract from the worth of the intention. “The addition of the performance of the deed adds nothing to alter the moral value” (14.20–21). Abelard rejects in general the notion that goodness need be summative: if *x* is good and *y* is good, the goodness of the conjunction of *x* and *y* need not be the sum of their individual goodnesses; indeed, a conjunction which is good need not be made up of individual good parts (46.17–48.30). Since deeds are only good paronymously, *i. e.* they are only called ‘good’ in virtue of stemming from a good intention, it is clear that the goodnesses of the intention and of the deed are not summative (52.4–15).

If deeds are indifferent, then their performance or nonperformance is equally indifferent. But an objection on Abelard’s own terms might be raised: what if the performance of a deed causes people to have certain reactions? For example, the intention to commit adultery is evil; even given that its performance is indifferent, doesn’t the pleasure the adulterer takes

<sup>17</sup> See also Abelard’s Epistle 7 to Heloise, edited by T. P. McLaughlin, “Abelard’s Rule for Religious Women,” *Medieval Studies* 18 (1956) 256, for a similar principle. Abelard states the principle in the following way in his response to q. 24 of the *Problemata Heloissae*, as given in J.-P. Migne, *Patrologia latina* 178, 710B: “The Lord, bringing all things back to the intention, assesses men to be condemned on the basis of things in the heart rather than things apparent in the deed, and He judges that the soul is only tainted on the basis of things in it.”

<sup>18</sup> See Diogenes Laertius, *Lives of the Eminent Philosophers* 7.104.

in the act add to its evil, or add a new evil? Abelard replies that the feeling of pleasure does not add to the evil involved, making his point with a lively example (20.15–18):

For example, if someone forces a monk (*religiosus*) to lie bound in chains between two women, and by the softness of the bed and the touch of the women beside him he is brought to pleasure, but not to consent, who may presume to call this pleasure (which nature makes necessary) a fault?

We are constructed in such a way, Abelard tells us, that the feeling of pleasure is inevitable in certain situations: sexual intercourse, eating delicious fruit, and the like. If sexual pleasure in marriage is not sinful, then the pleasure itself, inside or outside of marriage, is not sinful; if it is sinful, then marriage cannot sanctify it—and if the conclusion be drawn that such acts should be performed wholly without pleasure, then Abelard remarks they cannot be done at all, and it was unreasonable (of God) to permit them only in a way in which they cannot be performed (20.1–6).<sup>19</sup> Thus the performance or nonperformance of a deed is indifferent, as are its associated consequences, including the pleasure it may involve.

Three of the four candidates for the determinant of moral worth—the deed performed, the agent's desires, the character of the agent—have been eliminated. Therefore, Abelard concludes, intention must be the determinant of moral worth. So stated, there is an obvious gap in his argument. Why can't the agent's intention *combined* with one of the other factors determine moral worth? Abelard must be relying on some principle such as "if *x* is only effective in the presence of *y*, and *y* can be effective without *x*, then *x* is irrelevant."<sup>20</sup> Abelard's positive case for intentionalism will justify passing over combinations.

## 2. The Case for Intentionalism

### 2.1. Why Intentions are Morally Relevant

Abelard's first positive argument on behalf of the agent's intention as the key ingredient in moral worth is that there is no other way to make coer-

<sup>19</sup> Abelard does not consider cases in which the reaction is not necessitated by nature but only by a person's character (or not at all). Presumably such cases are irrelevant, since the important point is not to *intend* to tempt another or cause another to sin, but more is needed here by way of explanation and defense.

<sup>20</sup> A weaker principle would be: "If *x* is only effective in the presence of *y*, then *y* is the proper or primary source of the effected results." But this principle seem to be false, since it treats all *sine qua non* conditions as genuine causes.

cion and ignorance morally relevant. Ignorance, as a cognitive feature of the agent, seems utterly removed from any deed-based morality, and coercion seems equally removed as well. In fact, with regard to ignorance, Abelard points out that simple ignorance is not in question but rather *negligence* is the central notion. If the ignorance is what Abelard calls 'invincible,' then the agent is not negligent in being ignorant, and so cannot be blamed. Abelard only mentions one other case of non-negligent ignorance, namely what cannot be foreseen (66.19), and clearly there is much more of a story to be told here. But the cognitive failure of the agent—no matter whether it involves ignorance or negligence—and the force of coercion *only make sense*, he argues, if moral worth is a function of the agent's intentions.

Abelard, typically, takes an extreme case to make his point. He argues that the crucifiers of Christ were not evil in crucifying Jesus. (This example, and others like it, got Abelard into trouble with the authorities, and it isn't hard to see why.) The unbelief of Christ's crucifiers does not suffice to make their intentions evil. Indeed, Abelard claims that they would have sinned if they had thought that crucifying Christ was required and did *not* crucify him (66.30–34):

Those who persecuted Christ or his disciples, believing that they should be persecuted, 'sinned in deed,' but they would have committed a heavier sin in fact if they had spared Him against their own consciences.

From this example Abelard draws two consequences. First, the only evil is to act against conscience. Now 'conscience,' for Abelard, is the faculty by which what is done is estimated to be pleasing or displeasing to God.<sup>21</sup> Second, he offers a criterion for the goodness of intentions (55.20–23):

An intention should not be called 'good' because it seems good, but because in addition it is just as it is assessed to be—that is, when, believing that what one intends is pleasing to God, one is not deceived in one's own assessment.

To formulate Abelard's criterion briefly:

An intention is *good* if and only if the intention is believed to and in fact does conform to God's will

Any intention which is believed not to conform to God's will is automatically evil, even if in fact it does conform to God's will. If I intend something God

<sup>21</sup> Abelard takes belief in God to be a rock-bottom fact, though he recognizes a wide variety of opinions about the nature of God: Christian, Jewish, Moslem, Pagan—presumably he would have found atheism incoherent. However, as we shall see, the reference to God will turn out to be eliminable in favor of a formal requirement on the structure of intention.

would approve but I mistakenly believe that God would not approve, then my intention is evil. An intention which is believed to and does conform to God's will is unqualifiedly good. An intention that is believed to but in fact does not conform to God's will may be good or evil; this is where questions of negligence arise. This can be summarized as follows:

	Conforms	Doesn't Conform
Believed to conform:	<i>good</i>	[ <i>depends on the case</i> ]
Believed to not conform:	<i>evil</i>	<i>evil</i>

For this moral schema to be applicable, individuals clearly have to be able to know God's will, or to be well-situated with regard to excuses. But before explaining Abelard's views about God's will, the topic of §2.2, we have to consider two obvious objections to his intentionalism. First, how is it possible to commit evil voluntarily—to want to be worthy of damnation? Second, since intentions are not accessible to anyone other than the agent, doesn't Abelard's view entail that it is impossible to make ethical judgments?

With regard to the first objection, Abelard has a two-fold answer. First, it is clear that we often want to perform the deed and at the same time do not want to suffer the punishment. A man wants to have sexual intercourse with a woman, but not to commit adultery; he would prefer it if she were unmarried (16.18–18). Second, it is clear that we sometimes “want what we by no means want to want”: our bodies react with pleasure and desire independently of our wills (see the discussion of pleasure in §1.3 above). If we act on such desires, then our action is done ‘of’ will, as Abelard calls it, though not voluntarily. There is nothing evil in desire: there is only evil in acting on desire, and this is compatible with having contrary desires.

With regard to the second objection, Abelard grants the premiss that others cannot know the agent's intentions.<sup>22</sup> However, Abelard does not take ethical judgment to be an interesting question. God is the only one with a right to pass judgment. Yet this fact doesn't prevent us from enforcing canons of human justice, because, Abelard holds, human justice has primarily an exemplary and deterrent function. In fact, Abelard argues, it can even be just to punish an agent we strongly believe had no evil intention. He cites two cases. First, a woman accidentally smothers her baby while trying to keep it warm at night, and is overcome with grief. Abelard maintains that we should punish her for the beneficial example her punishment

<sup>22</sup> God, however, has access to internal mental states, and so there can be a Final Judgment. The mental is only private *per accidens*, as it were, for mediaeval philosophers.

may have on others: it may make other poor mothers more careful not to accidentally smother their babies while trying to keep them warm. Second, a judge may have an excellent but legally impermissible piece of evidence that two witnesses are perjuring themselves, and so be judicially forced to rule that the accused, whom the judge believes to be innocent, is guilty. Human justice may with propriety ignore questions of intention.<sup>23</sup> Since there is divine justice, ethical notions are not an idle wheel—nor could they be, even on Abelard's understanding of human justice, since they are the means by which we determine which intentions to promote or discourage when we punish people as examples or in order to deter others.

## 2.2. A Mediaeval Categorical Imperative

Abelard has been arguing against (P) to justify, at least in part, the adoption of (P\*). There are other reasons, to be sure; it is precisely because (P\*) rather than (P) holds that Abelard can explain the Biblical story of Abraham and Isaac, that is, to hold that it may be good to intend what would not be good to perform. But the central problem with (P\*) hasn't been addressed so far. How can the goodness of an intention be characterized independently of the goodness of its associated action? Isn't (P\*) parasitic on being able to identify the goodness of the deed independently of the agent's intentions?

Both (P) and (P\*) suppose that there is an independent way to tell what is good: we need to know what deeds are good to bring about for (P), and we need to know what intentions are good to have for (P\*). Now as suggested in §2.1, the appropriate knowledge to have for (P\*) is whether an action conforms to the will of God. Since Abelard rejects desire as appropriately moral, he correctly deduces that an agent who acts out of *fear* of God is not acting morally. That is, an agent is not moral if she subordinates her will to God's will only out of fear of damnation, which Abelard calls "God's justice." Rather, an agent should subordinate her will to God's will solely for the love of God, which Abelard calls "God's mercy." This readily suggests a moral commandment that is binding on all agents, a mediaeval 'categorical imperative':

Always act such that the intention according to which you act has the form, or could correctly be represented as having the form, "to intend to  $\varphi$ -for-the-sake-of-God's-mercy."

<sup>23</sup> Actually, much of the law is given over to determining the intentions of the agent at the time of performance. But this is to grant Abelard's point that intentions are relevant to morality, although to try to usurp divine privilege.

No intention having this form can be evil; whether any such intention is good depends on having the correct beliefs about God's will and His mercy, as noted above. Equally, Kant's categorical imperative is usually construed as a negative test; it only prohibits certain actions (those whose maxims cannot have the form of a universal law). This allows agents to perform deeds they do not desire—in Abelard's terminology, mentioned in §1.1 above, to 'endure' performing  $\varphi$ —and subordinate their will to God's will. The subordination of one's own will to God's will is for Abelard the summit of morality.

There might seem to be a disanalogy with Kant's categorical imperative. Kant argues that his categorical imperative is binding on all rational agents according to the nature of reason as such, whereas Abelard's principle seems to depend on knowledge of God's will, which is not obviously cognitively accessible to every rational being, since it may depend on particular historical knowledge of particular historical events, such as the coming of Christ and the new revelation. Nor is Abelard's principle 'formal' the way Kant argues the categorical imperative must be.

But what is God's will? Abelard has a surprisingly modern answer. It is the fulfillment of the Natural Law, which is cognitively open to all rational beings, without any special revelation. He reiterates this point in a variety of works, though not, interestingly, in the *Ethics* (*Theologia christiana* II.44):

If we carefully consider the moral precepts of the Gospels we will find nothing more than a reformulation of the Natural Law, which the [pagan] philosophers clearly followed.

God's will, then, as embodied in the new revelation, has the exact content of the Natural Law. But the Natural Law is by definition accessible to all (*Dialogus* II. 2220–2224):

Natural Law is what reason, which is naturally present in every person and so remains permanent in all, moves us to perform: worship God, love parents, punish evildoers.

It might be wondered how the characterization of the Natural Law in terms of specific actions (*e. g.* loving one's parents) is accessible to all. But Abelard describes the content of the Natural Law in a purely formal way (*Commentary on "Romans" ad 2:13*):

The words of the Natural Law are those which enjoin charity to God and to one's neighbor, that is, 'do not do unto others as you would not be done unto' and 'do unto others as you would be done unto' [*Matthew* 7: 12].

Hence Abelard's principle can be reformulated in the following way:

Always act such that the intention according to which you act has the form, or could correctly be represented as having the form, “to intend to  $\varphi$  for the sake of doing unto others as you would be done by or not doing unto others as you would not be done by.”

This reformulation, I think, has as much of a claim as Kant's categorical imperative to be binding on all rational agents from the structure of reason and to be strictly formal. Abelard offers a generalizability-test for the goodness of intentions, much as Kant offers a universalizability-test for maxims. But the parallels between Abelard's and Kant's versions of intentionalism don't rest only on this reconstruction of Abelard's views about the goodness of intentions. Their arguments for their respective versions of intentionalism are astonishingly close, even on points of detail, as a more detailed comparison will establish.

### 3. Kant's Metaphysical Morals

The first section of Kant's *Groundwork of the Metaphysics of Morals*, entitled “transition from the common rational knowledge of morals to the philosophical,” is designed to uncover the intentionalist core of traditional (Christian) morality.<sup>24</sup> First, Kant asserts, notoriously, that the only thing that is good without qualification is a good will. He argues, in two stages, that desires are irrelevant to moral worth. First, because any action stemming from (mere) inclination is irrelevant to morality; those who do what they please might “deserve praise and encouragement, but no esteem” (398). Second, imagine a man whose mind is “clouded by sorrow,” who manages to “tear himself, unsolicited by inclination [desire], out of this dead insensibility and to perform this action only from duty and without inclination”—and then his deed begins to have moral worth. Finally, if a man is by temperament cold and indifferent, then his performance of such deeds is even more to be esteemed, since they have moral worth and are even contrary to his desires (398). Desire is not the determinant of moral worth.

Nor is moral worth a matter of the character of the agent. Moderation and other dispositions can be found in both good and evil men, and “the coolness of a villain makes him not only far more dangerous but also more directly abominable” (394). Furthermore, moral worth is not to be found in deeds. Kant notes, first, that a good will which is in fact unable to ever accomplish any of the purposes for which it strives would nonetheless “sparkle like a jewel in its own right” (394). His point is that deeds are in

<sup>24</sup> All references to Kant's arguments, unless otherwise noted, are to his *Groundwork of the Metaphysics of Morals*.

part a function of moral luck: which deeds one is able to perform, and the consequences they have, is in part dependent on contingent circumstances beyond the agent's control—and as such have no place in morality. Second, Kant adduces the example of the grocer who would gladly cheat his customers but he does not do it for fear of being discovered. The deeds are extensionally identical to those performed from a sense of duty, but they lack all moral value. Indeed, the “first proposition of morality” for which Kant argues is:

A deed has moral worth only if it is performed from a sense of duty. This first principle is of a piece with Abelard's claim that deeds are in themselves indifferent, having value only in that they stem from a good intention. Both Abelard and Kant adopt this piece of classical Stoicism. Kant's “second principle of morality” is astonishingly reminiscent of Abelard's central claim (399): A deed performed from a sense of duty does not have its moral worth in the purpose which is to be achieved through it but in the maxim by which it is determined.

In a footnote, Kant informs us that a ‘maxim’ is the “subjective principle of volition.” A natural way to read this would be to take a maxim to be an intention, which also seems to qualify as a subjective principle of volition.<sup>25</sup> Therefore, Kant is asserting that the sole determinant of the moral worth of an action is the intention according to which it is performed—not the deed itself or its purposes (or, presumably, its consequences).

From these two propositions, Kant argues that moral principles are by definition formal: they “condition” the will “without reference to any expected result.” Kant concludes that the only way this can occur is “through the universal conformity of the will to law as such” (402). In less obscure terminology, Kant is claiming that the maxim or intention has to be universalizable without contradiction. There is much more to be said about universalization, and even more about the relevant sense of ‘contradiction,’ but for our purposes the important point to note is that Kant argues that intentions have moral worth solely in virtue of their formal structure. Abelard endorses the same claim, but locates the formal structure of the intention in conformity to the Golden Rule (or to  $\varphi$  for-the-sake-of-God's-mercy). Kant builds in the cognitive accessibility of his “Supreme Principle of Morality”

<sup>25</sup> Kant holds that maxims are given, not formulated, and hence not in our control: we ‘discover’ the maxims on which our actions are based. (This in part gets around problems with logical fiddling to pass the universalizability test.) Abelard's account of intentions is much less clear. He seems to think that intentions are at least sometimes consciously devised and adopted by us. If so, this is a point of deep disagreement between Abelard and Kant.



by claiming to have derived it from the structure of rationality as such. Abelard, too, claims that the Natural Law is naturally implanted in all agents and accessible by the exercise of reason. Equally, Kant and Abelard each have to provide an account of how happiness is linked up with morally correct action (or dispositions to morally correct action, *i. e.* virtue). That is the project of Kant's second *Critique* and of Abelard's *Dialogus*, discussed in §4.

The comparison of Abelardian and Kantian intentionalism suggests that much of what is valuable in Kant's ethics can be divorced from his idiosyncratic metaphysics: Abelard holds a substantively similar theory without saying anything about the noumenal and phenomenal realms. Again, the comparison suggests that the value of Kantian ethics is not so much to be found in its (supposed) prescription of absolute rules of conduct, but rather in its effort to define what permissible intentions an agent may have. Finally, Kant's theory can be supported by an appeal to Abelard's arguments, which are richer and more detailed than any Kant puts forth.

#### 4. Stoicism and the Supreme Good

Just as Abelard's *Ethics* is appropriately paired with Kant's *Groundwork of the Metaphysics of Morals*, so too Abelard's *Dialogue Among a Christian, a Jew, and a Philosopher* is appropriately paired with Kant's *Critique of Practical Reason*: each explores the connections among the supreme good, morally correct action (in accordance with duties for Kant and with the virtues for Abelard), and happiness. In addition to this thematic unity, the two works are united by a deeper underlying project, namely baptizing classical Stoic metaethics. Abelard and Kant revise the Stoic equation of virtue with happiness to allow conceptual space for the moral role played by God and the Afterlife—constituent elements in the supreme good. But they pursue this common project in radically different ways.

For Kant, to act virtuously—that is, action in accordance with duty—depends on the fundamental practical postulate of a divine providence that will guarantee the reward of (present) virtue (in this life) with (future) happiness (in the Afterlife). Otherwise, Kant argues, practical reason would be incoherent, since it would require the agent to seek happiness and also to follow duty, although they pull in different directions. Now there is an important sense in which Kant doesn't engage the Stoic theory on its merits. For he explicitly identifies the link between virtue and happiness as a 'fundamental practical postulate': a thesis endorsed by the will and applied beyond the scope of pure reason. That is, the link is unknown and unknowable, not established by argument but deduced from the nature of

practical reason itself, a fact overlooked by the Stoics.<sup>26</sup>

Abelard, in contrast, makes the conflict between Stoicism and Christianity one of the central themes of his *Dialogue*, it being the pivot around which the conversation of the Philosopher with the Christian revolves. Yet for Abelard's Philosopher and Christian, the existence of God and the Afterlife is common rational ground, as is the existence of God; these are reasoned positions, not 'postulates'. The interest of Abelard's discussion is located in the disagreement between the Philosopher and the Christian over the relation of the supreme good to virtue and happiness, and in particular whether the Afterlife is a state intrinsically tied to happiness. The Christian maintains that it is, the Philosopher that it is not. The Philosopher's capitulation on this point is Abelard's reply to classical Stoicism.

Abelard's reply, however, is unsatisfying. The Philosopher capitulates to the Christian by giving up one of the central tenets of Stoicism, namely the distinction between moral and nonmoral goods, but he does so for no good reason, and Abelard's failure to convince us on this point is perhaps as ultimately unsatisfying as Kant's move to a rationally indefensible practical postulate. A closer look at the course of the debate at this point in the *Dialogus* is in order.

The Philosopher is arguing that virtue entails happiness, and hence there is no need of an Afterlife since a virtuous person remains in the same condition whether dead or alive. The Christian claims that the Afterlife must be something better than any condition attainable in this life, for otherwise the Afterlife "was mistakenly proposed as a reward if it is not better or more pleasant than the present life" (*ll.* 1640–1643). Now one would expect the Christian to go on to argue that the Afterlife is in fact better than the present life. Instead, the Christian merely points out that such an argument is necessary. The Philosopher supplies the argument, committing a form of dialectical suicide by arguing himself out of his own position, and it turns out that the Afterlife is 'better' than our present life because it doesn't rain (*ll.* 1652–1660):

However, as long as something prevents our will or is absent from it, there isn't any true blessedness. And this surely is always the case so long as one is living here, and the soul, burdened with the weight of its earthly body and shut up in it as though it were a prison, does not enjoy genuine freedom. For who doesn't sometimes desire heat when it's too cold (or vice versa), or good weather when

<sup>26</sup> This is by and large the story Kant tells in his famous footnote on the Stoics in the *Critique of Practical Reason* II.ii.5 (Akad. 127).

rain is burdensome, or often want more food or clothes than he has? And there are countless other matters, if we don't resist the evident truth, that befall us when we don't want them or are denied when we do want them.

The Philosopher concludes that the Afterlife, where even such petty frustrations of the will are lacking, is indeed an improvement on anything available in this life, and further concedes to the Christian that present virtuous life prepares us for the 'reward' of such an Afterlife. Thus does Abelard try to baptize classical Stoic metaethics.

Yet the Philosopher, in this crucial argument, is appealing to *nonmoral* goods—no rain, no lack of food, and the like. But no amount of nonmoral goods can secure the claim that the Afterlife is (morally) better.<sup>27</sup> Classical Stoicism drew an important distinction here, one that Abelard is ignoring without benefit of argument. The goods that the Philosopher is referring to are indifferent, as we have seen demonstrated at length in §1.3, but they are nevertheless goods to be *preferred* (προηγμένα) rather than goods to be *rejected* (ἀποπροηγμένα).<sup>28</sup> Such goods are purely nonmoral and, though it is better to have them than to lack them, this is not a matter of moral goodness. The Philosopher is only entitled to conclude that the Afterlife is preferable, and perhaps even preferable for its own sake (δι' αὐτὰ: DL 7.107), not that it is morally superior.<sup>29</sup> His capitulation to the Christian is a mistake.

### Conclusion

Abelard's attempt to baptize Stoic metaethics, then, is unsatisfying in the end. But this failure doesn't indict his normative position, his intentionalist ethics, any more than we have to accept Kant's postulates of practical reason to be Kantians in moral philosophy. That's fortunate, since there are good reasons to take intentionalism seriously as an ethical theory, whether in its Stoic, Abelardian, or Kantian versions, having to do with the distinction between events and actions and why morality concerns the latter but not the former.

<sup>27</sup> We could claim that moral goodness is closely linked to the possession or maximization of nonmoral goods, as, say, hedonistic utilitarians do. But Abelard doesn't do this, for the reasons given in §1 above.

<sup>28</sup> See Diogenes Laertius 7.104–6 for the Stoic account of preferred and rejected goods.

<sup>29</sup> The Christian carefully says that the Afterlife must be "better or more pleasant than the present life" (*quae nisi melior uita praesente sit aut magis placeat*: ll. 1640–1641), but this isn't enough for the argument.

Earthquakes are events, often disastrous events, that cause pain and suffering, misery and death. Yet they aren't *morally* evil. They aren't the sort of thing that can be morally good or evil. Why not? The straightforward answer, which in this case I think is the correct one, is that an earthquake is simply an event—a mere physical process, if you like. They are events but not actions. Only some events are actions, and they are those events not merely in which a person is present (people may be, unhappily, present when an earthquake takes place), but those events that occur as the result of *intentional activity*. If moral goodness and evil is to be a special property of human behavior and not to apply to earthquakes, the difference between actions and events must be a morally relevant difference. Therefore, intentional action, which is what distinguishes a (mere) event from an action, must be crucial to the enterprise of morality.<sup>30</sup>

Intentionalist ethics takes this point seriously, and tries to account for it by identifying the capacity for intentional action as the distinctive feature of moral agency. Whether any given version of intentionalism succeeds as a moral theory is another question, to be sure, but one that shouldn't blind us to the genuine philosophical issues that such theories address. Abelard's version of intentionalism, in particular, is subtle and sophisticated, and should earn him philosophical respect—not merely in the history of ethics, but in contemporary discussions of ethical theory, since his intentionalist ethics is a genuine alternative to most positions under discussion in modern moral philosophy.

<sup>30</sup> This fact is often passed over in contemporary moral philosophy. Utilitarianism, for instance, is concerned with which states of affairs ought to obtain, and traditionally favors those states of affairs in which nonmoral good is maximized for the greatest number. Now given that intentions are a *sine qua non* for moral assessment, it is impossible to characterize the goodness and badness of actions independently of the intentions they embody. But that is precisely what utilitarianism does, by pointing to (for example) pain and suffering—nonmoral goods—independent of their origin. Yet there is no way to link these nonmoral goods to morality without having recourse to the agent's intentions. Contemporary 'deed-centered' theories of morality all seem to share this feature.