1. The double ambivalence of human nature

1. THE EVOLUTIONARY VIEW OF HUMAN NATURE: NEITHER PESSIMISTIC NOR OPTIMISTIC

In the discussions concerning evolutionary theory's implications for the understanding of human nature, one can distinguish two extreme positions (cf. de Waal 2006). On the one hand, one may argue that human nature is amoral, lacking in moral dispositions; on the other, it may be considered that we come to the world with well-developed moral dispositions, and lacking in immoral dispositions.

Perhaps the most famous defence of the first, pessimistic, view was provided by Thomas Huxley in his 1894 essay *Evolution and Ethics*. He claimed that the characteristic feature of man is 'self-assertion, the unscrupulous seizing upon all that can be grasped, the tenacious holding of all that can be kept'; that man's behaviour 'is adjusted to cosmic strife [and] manifests ruthless and ferocious destructiveness when his anger is roused by opposition' (Huxley 1995 (1894): 129). This pessimistic picture of man was strictly connected with Huxley's gloomy view of the world; as he wrote, 'this world is full of pain and sorrow; grief and evil fall, like rain, upon both the just and unjust … Brought before the tribunal of ethics, the cosmos might well seem to stand condemned. The conscience of man revolted against the moral indifference of nature' (Huxley 1995 (1894): 130). An even stronger version of this view was defended by George C. Williams, who asserted that natural selection is

> a process of maximizing short-sighted selfishness … The survival of one organism is possible only at the great cost to others … Huxley did not go far enough in his condemnation of the evolutionary process. We now have a far deeper understanding of evolution than Huxley had, and far more information on what this process has wrought. Huxley, for instance, understood nothing of forces favouring the development of nepotism or of parent–offspring and other sorts of conflict in nature. He knew little of the prevalence and violence of infanticide and cannibalism. (Williams 1995: 318)

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

The other extreme view – the optimistic one – assumes that biological evolution has endowed us with a number of prosocial instincts which directly lead to moral behaviour. Darwin (1871: 67) distinguished four such tendencies or instincts: 'taking pleasure in the society of one's fellows'; 'feeling sympathy for them'; 'performing various services for them'; and 'regarding others' approbation and disapprobation'. One cannot say, however, that Darwin assumed the optimistic view: he was deeply aware of the possibility of conflicts between social instincts and other instincts, for example the instincts of self-preservation, hunger, lust and vengeance (and also of the morally imperfect or incomplete character of the social instincts themselves). Nonetheless, his insight regarding the presence of prosocial instincts in human nature could have led other evolutionary thinkers, such as the famous anarchist philosopher Peter Kropotkin, to adopt the optimistic view.

It must be admitted, though, that of these two extreme views, the optimistic one was adopted much less frequently than the pessimistic. But both views seem to be implausible: they are onesided, failing to give justice to the complexity of human nature, its multilayered character. I shall argue that our 'evolutionary' nature can be most aptly characterized by the term 'double ambivalence'.

The view of human nature as doubly ambivalent is neither pessimistic nor optimistic; rather, it is a combination of both. This view is based on two claims. The first is that human nature comprises immoral tendencies (egoism, envy, aggression, self-deception), neutral tendencies (some emotions, hierarchical propensities, the desire for self-transcendence) and moral tendencies (empathy, kin and reciprocal altruism, tribalism). This is the *first ambivalence*: human nature is not monolithic in its moral fibre (though I do not insist that one could not find other examples of immoral, neutral or moral tendencies apart from the ones mentioned above and analysed in further parts of this chapter). The *second ambivalence*, more interesting than the first one, is that the moral tendencies themselves are morally imperfect (far from what constitutes or gives rise to what I call 'genuine ethics'), and thereby morally ambivalent. It is therefore not true (*contra* pessimists) that our nature does not determine any kind of morality: our nature comprises a number of moral tendencies whose 'correlates' are rules of what I call 'evolutionary ethics'. However, the morality determined by these tendencies is (*contra* optimists) very imperfect: it may lead to immoral behaviour, which I have called 'secondary evil'. Accordingly, human nature, *even in its better part*, constitutes our 'gravity' (*pesanteur*), in Simone Weil's sense (1988: 132) – that is, the force of the natural impulses of the soul, which we should oppose; for example, the impulse to seek moral equilibrium by taking

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

revenge or by awaiting reward (cf. Weil 1988: 132). As Weil aptly remarks, it is not by chance that we speak about 'falling into' only in the context of evil; we cannot 'fall into' goodness, since goodness is something 'above' us, towards which we have to soar, overcoming our natural impulses, which drag us lower.

There is, however, an important optimistic element of the view of human nature defended in this book: the view implies that human nature does not contain *demonically* amoral tendencies, that is, human beings do not possess evolutionarily shaped tendencies to commit extreme evil – absolute, sadistic, or 'unintelligible'. In fact, the nonexistence of extremely evil proclivities is a simple corollary of the logic of biological evolution. A propensity can be a biological adaptation only if it increased our ancestors' biological fitness (the chances of survival and reproduction). But this biological goal, which each evolutionary adaptation must, by definition, serve, is, of course, ethically acceptable. Consequently, even immoral tendencies which were shaped by natural selection are not *totally* immoral, because they (ultimately) serve an ethically acceptable goal. One may query, however, how to reconcile this (optimistic) conviction with all too many testimonies to human cruelty: that cruelty *for its own sake* may be perpetrated by humans is a sad fact (cf. Zdziechowski 1993, or the memorable chapter about *violenza inutile* (useless violence) from Primo Levi's famous 1986 book *I sommersi e i salvati*, published two years later in English as *The Drowned and the Saved*). The simple answer to this question is that just as human beings are capable of going *above* their 'evolutionary ethics' (determined by their evolutionarily shaped moral tendencies) and reaching the stage of 'genuine ethics', so they may fall *below* this level. This is due to the fact that human beings, thanks to their reasoning capacities, may, so to say, free themselves (to a certain but considerable degree) from their biological nature not only in the right direction but also in the wrong one.

In the remainder of this chapter I shall develop at greater length the view of human nature as doubly ambivalent.

2. IMMORAL TENDENCIES

2.1 Egoism

In moral–psychological analyses, egoism is usually assumed to be a primitive, 'unanalysable' motive, that is, one which does not flow from some other more basic psychological phenomena. However, this widely accepted and apparently self-evident assumption seems to be mistaken. I

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

shall argue that: (1) egoism is not a primitive motive but, rather, a manifestation of some more basic psychological phenomena; (2) one can distinguish three different forms of egoism depending on its psychological basis, viz., cognitively based, *hybris*-based and instinct-based; (3) two of these three forms of egoism (viz. cognitively based and instinct-based) can be plausibly viewed as a product of biological evolution.

In my analysis I shall assume a commonsense definition of egoism as an agent's tendency to pursue his own interests to an exceedingly high degree, that is, without duly respecting the other agents' interests. This definition presupposes that pursuing one's own interest is not in itself morally wrong – it becomes morally wrong only when it is done 'in an exceedingly high degree', and the degree of the morally legitimate – nonegoistic – pursuit of one's interest is determined by morality, which says how much respect is due to other people's interests in the course of pursuing one's own interest. Egoism can be defined equivalently in at least two different ways. Schopenhauer defined egoism as 'an urge to one's being and well-being' (*der Drang zum Dasein und Wohlsein*) (Schopenhauer 2009: 662; 2010: 196). This definition requires a slight correction, since, as already noted, not all types of care for oneself can be viewed as egoism. Egoism is an *excessive* 'urge to one's being and well-being' (an insufficient urge of this kind is abnegation). Egoism can also be defined by modifying the Evangelical precept 'love thy neighbour as thyself' (Mark 12:31; Matthew 22:39; Luke 10:27), which arguably states the conditions of the morally right conduct towards other people. Now, one can say that a person is egoist if and only if he loves himself more than his 'neighbour'. Thus, we have three arguably equivalent definitions of egoism: as an agent's tendency to pursue one's interests without duly respecting the other agents' interests, as an excessive 'urge to one's being and well-being' and as loving oneself more than one's 'neighbour'.

My basic claim is that egoism may flow from three different psychological sources: cognitive bias of egocentrism, *hybris*, and biological instincts of self-preservation and reproduction. Accordingly, one can distinguish three forms of egoism: cognitively based, *hybris*-based and instinct-based. I shall now present these three forms in more detail.

Cognitively based egoism flows from overestimating the 'reality' of oneself as compared with the 'reality' of others (that is, from egocentrism). For an agent who manifests this form of egoism, other people are substantially 'less real' than herself – or even 'unreal'. This conviction leads to a diminished capacity to consider other people's perspectives and to the tendency to interpret events with excessive reference to

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

oneself. This account of egocentrism is somewhat different from, though related to, the ordinary type, according to which egocentrism is an excessive preoccupation with oneself, which may (but does not have to) result in not paying sufficient attention to or even completely neglecting the fact that other human beings have their own 'inner worlds'. Thus, on the ordinary account, an egocentric person does not necessarily downplay the 'reality' of other people (she is just excessively preoccupied with herself), whereas on my account an egocentric person downplays the 'reality' of other people and, as a result, may (but does not have to) become excessively preoccupied with himself. Egocentrism (on my account) seems to be at least partly caused by the fact that we have direct access to our ego, to our mental life, and lack direct access to other people's egos, to their mental lives; other people's mental states (beliefs, emotions, attitudes) can only be known indirectly. This indirect character of our cognition of other people's mental states is particularly salient when such cognition takes the form of inferring (by analogical reasoning) those other people's mental states from their behaviour. But it occurs also when the cognition proceeds via *Einfühlung* (a kind of psychological insight into another person's mental state): even this form of cognition does not possess the 'directness' which is typical for self-knowledge. It is clear that the fact that we have privileged access to our own mental life does not have to lead to egoism, although it seems always to lead to the conviction that other people are somewhat less real than ourselves. We are all to some extent egocentric. This privileged access leads to egoism only if egocentrism assumes (for reasons to be ascertained by empirical psychology) the form of *strong* egocentrism, that is, if it generates in the agent a conviction that other people are *substantially* less real than himself or even unreal – that, figuratively speaking, the ontological status of other people resembles that of 'shadows' (regarded more as objects than as real persons). Thus, it is strong egocentrism, not egocentrism *simpliciter*, that underlies a cognitively based egoism.

Valuable insights into the phenomenon of egocentrism have been provided by social and cognitive psychologists who have identified and analysed in depth many of its manifestations, such as, for example, a cognitive bias called the 'spotlight effect' (cf. Gilovich, Medvec, Savitsky 2000). The spotlight effect consists in people's belief that they are noticed more than is really the case. They believe that they are in a 'social spotlight', overestimating their own importance to other people. This phenomenon is obviously egocentric in nature: since people are at the centre of their own world, they tend to believe that they are also at the centre – or close to the centre – of other people's worlds. In the case of the spotlight effect, overestimation of one's own 'reality' (which is

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

characteristic of each manifestation of egocentrism) does not therefore lead to discounting the 'reality' of other people, but rather to overestimating the 'reality of oneself' in the consciousness of other people. The spotlight effect may be more or less acute for different people. It seems that those who manifest it particularly strongly are more likely than others to be subject to strong egocentrism, that is, to discounting the reality of other people, and in consequence to cognitively based egoism. I have discussed here only one egocentric 'effect' discovered by social and cognitive psychologists, but there are others too (some of them strictly connected with the spotlight effect), such as 'the illusion of transparency' (overestimation of the degree to which one's mental state is known by others); 'self-as-target bias' (the agent's belief that the course of events in the world is particularly 'inimical' to her own plans and intentions – more so than to the plans and intentions of other people); 'false consensus effect' (overestimation of the degree to which other people share one's beliefs and emotions); and the opposite effect, 'false uniqueness effect' (underestimation of the degree to which other people share one's beliefs and emotions).

By way of a historical digression, one may note that Schopenhauer (cf. 2009: 630) hinted at the claim that egoism may be a result of our having direct access only to our own mental life, and thereby seeing other people as 'representations'. He developed this insight in the spirit of the Upanishads, suggesting that the correct cognitive stance consists not in recognizing that other persons are real but in negating the reality of one's own ego. Cognitively based egoism can therefore be overcome in two different ways. The traditional ('Western') 'therapy' to work against it consists in recognizing that other people have the same – *full* – reality as oneself, whereas the Upanishads therapy (the 'Eastern' type) consists in recognizing that other people have the same – *none* – reality as oneself, and that the only reality is that of the 'cosmic ego' (Atman–Brahman). It is a matter of dispute whether the practical consequences of liberation from egoism are the same irrespective of which of these two anti-egoistic therapies one undergoes. One could argue that acceptance of the belief (endorsed by the Upanishads and Schopenhauer) that the reality of an individual ego is apparent may discourage one from performing acts of benevolence, because if other people's egos are unreal, it is unclear why one should care at all for the wellbeing of those other people. The fact that one's own ego is also unreal is not a plausible answer; it begs the question. The only plausible answer could be that all people are 'immersed' in the nonindividual ego (Atman–Brahman). But, following Max Scheler, one could ask whether beneficial acts thus motivated are not in fact a 'camouflaged egoism' (Scheler 1980: 78): if one helps

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

another knowing that the other is in fact oneself (in accordance with the famous teaching of the Upanishads summarized in the phrase *Tat-Twam-Asi* or 'that art thou', that is, your individual ego hides at its depths – is identical with – the cosmic ego), then in helping others one in fact helps oneself. The above remarks also provide an answer to the question of whether cognitively based egoism is morally blameable. The answer is obviously in the affirmative, for the simple reason that this form of egoism is relatively easily avoidable: most people can overcome, by a relatively small mental effort, their impression of 'the lesser reality' of other people – in other words, they can easily overcome their egocentric proclivities or at least prevent them from turning into strong egocentrism and, consequently, into cognitively based egoism. The exception is to be made for those who suffer from some cognitive defect – some kind of a deficit of mental energy – that incapacitates them from overcoming their impression of the 'lesser reality' of other people. A precise description of this kind of defect belongs to the domain of psychopathology.

Instinct-based egoism is an excessive manifestation of biological instincts of self-preservation and reproduction. These instincts 'move' human beings to undertake actions serving the realization of their basic biological goals even if performing these actions means violating other people's morally justified interests. Of the three varieties of egoism, the instinct-based one is the most distinctly hedonistic (evolution 'has motivated' us to pursue our biological 'goals' through the promise of pleasure accompanying their realization). One can therefore say that a behavioural manifestation specific to this form of egoism consists in attaching undue importance to hedonistic values at the price of neglecting other – higher – values. We come to the world with this form of egoism present, and to overcome it is the fundamental purpose of education as well as of moral self-improvement. Clearly, these educational and self-improvement efforts are not always successful: instinct-based egoism becomes a dominant motive of action of many people. It is worth noting that it is this form of egoism that Immanuel Kant had in mind when he wrote about egoism (the tendency to prioritize self-interest over the interests of other people and over moral law) as the basic obstacle to our following moral rules, and thereby as a *conditio sine qua non* of the 'radical evil' (*das radikal Böse*) of human nature (cf. Kant 1986: 21–56). Thus, in Kant's view, egoism is at the root of human inability to unwaveringly follow moral principles (which is the essential content of his doctrine of 'radical evil', which I analyse in more detail at the end of this chapter).

Hybris-based egoism does not flow from the impression that other persons are substantially less real or unreal, but from the conviction that other persons are of lesser worth – that is, from *hybris*. *Hybris* is the

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

belief in one's own superiority over other persons and in one's having special rights and privileges flowing from this purported superiority. It therefore implies a blatant denial of a fundamental moral equality of human beings. *Hybris* manifests itself as self-confidence bordering on arrogance, insolence, overbearing pride, self-aggrandizement. It is the opposite of humility and reverence, the latter being

> the virtue that keeps human beings from trying to act like gods, does not let them forget that they are humans … An irreverent soul is arrogant and shameless, unable to feel awe in the face of things higher than itself. As a result, an irreverent soul is unable to feel respect for people it sees as lower than itself – ordinary people, prisoners, children. (Woodruff 2001: 3)

According to Aristotle, a particularly sharp manifestation of *hybris* is the deliberate humiliation of another person for no other reason than the pleasure involved. The cause of the pleasure of the *hybristic* man is that he finds himself greatly superior to others when ill-treating them. Thus, *hybris* manifests itself in an especially acute form when one gives offence neither for profit nor revenge, but simply because one delights in inflicting shame on others (cf. Elster 1999: 62–3, 173–8). Undoubtedly, *hybris*-based egoism is the most wrongful form of egoism, precisely because it flows from *hybris*, rightly regarded by the ancient Greeks as the most serious moral depravity and assumed in the Christian ethics (where it was called *superbia*) to be the most serious of human sins (and the root of *peccatum mortale* – mortal sin). One can find few circumstances to excuse this form of egoism: it is neither a result of natural egocentric tendencies (as cognitively based egoism is) nor a result of natural instinctive drives (as instinct-based egoism is), but a result of *hybris*, which is a malign, pathological product of human freedom.

Four additional comments are in order here.

First, since instinct-based egoism is not based on a wrongful feeling of superiority over other people, it is less morally reprehensible than *hybris*-based egoism, and since it is a result of the overabundance of instinctive drives (and the concomitant 'enslavement' to the pursuance of pleasure) which seem to be harder to control and overcome than egocentric proclivities, it also seems less morally reprehensible than cognitively based egoism.

Second, it is not clear whether the three varieties of egoism can be displayed at the same time in an agent's egoistic behaviour. Arguably, *hybris*-based egoism presupposes that other human beings are as real as oneself, so it cannot go in tandem with cognitively based egoism. Obviously, one person may manifest different forms of egoism at different times.

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

Third, egoism, if it is not unrestrained, seems to be biologically adaptive. It is generally in the interests of our 'genes' to place higher value on our own interests than on those of other people (unless these other people are our kin, with whom we share – to an extent which depends on the level of biological relatedness – our 'biological' interests). But, arguably, not all varieties of egoism are biologically adaptive. Egocentric cognitive biases appear to be biologically adaptive, since overestimation of the level of the reality of one's own person seems on the whole to increase our biological fitness. Similarly, strongly developed instincts of self-preservation and reproduction seem to be fitness-enhancing. However, it is dubious whether *hybris* or *superbia* – the belief in one's superiority over other persons and the resulting desire to subordinate other persons – is biologically adaptive; as will be argued in more detail in Chapter 2, it can be surmised that people who exhibited such tendencies were removed from ancestral communities, and their genes could not have been preserved in the gene pool. Therefore those who manifest *hybris* or *superbia* behave *below* their evolutionary nature.

Fourth, Adam Smith examined the question of whether people want to be moral or only want to be regarded as moral – or, in Smith's terms, whether they exhibit love or praiseworthiness, or only love of praise. Adherents of the theory of psychological egoists, such as La Rochefoucauld, denied that people have a true love of virtue; in these thinkers' view, people want only to conceal their vices and to pass for virtuous. Adam Smith disagreed, saying: 'Some splenetic philosophers, in judging human nature, have done as perverted individuals are apt to do in judging of the conduct of one another, and have imputed to the love of praise, or to what they call vanity, every action which ought to be ascribed to that of praiseworthiness' (Smith 2007: 159). Smith's view seems to be more consistent with our common experience and is assuredly not falsified by evolutionary theory: there is nothing in evolutionary theory which would forbid us to assume that human beings may be motivated by a desire for being worthy of moral praise and approval. Our egoism is just one of many motives we may have, not the only or even the dominant one.

2.2 Envy

The term 'envy' is ambiguous and can be applied to three quite distinct phenomena. What can be said about envy in general, with respect to its three forms, is that: (1) it is a three-place relation which embraces a subject (an envier), a rival (a party who is envied) and a good that the rival possesses and the envier does not (cf. D'Arms 2008); (2) it is a

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

manifestation of the human tendency to evaluate one's own situation in a comparative way, that is, by referring it to the situation of other people; accordingly, it attests to the fact that people are concerned not with their absolute level of goods but with the relative level (that is, compared with the standing of others). We are social beings, and our sociality manifests itself, among other things, in the tendency to compare ourselves with other people, to be concerned with our 'social self' – with how we are perceived by other people. If it had not been for this tendency, envy could not appear, though, arguably, this tendency is only a necessary, not a sufficient, condition of the appearance of this emotion.

I shall now present three basic forms of envy, viz. the malicious, the just and the benign (admiring).

Malicious envy contains two elements: (1) an unpleasant emotion (distress, pain, nuisance, and so on) felt by the subject at the thought that she does not possess the good and the rival does; (2) a desire that the rival lose the good – a desire which appears despite the fact that, in the envier's view, it is not unjust that the rival possesses the good. Malicious envy may take active or passive forms: it may move an agent to take steps leading to the envied person's deprivation of the good she possesses (active envy), or it may consist in passively counting on the envied person losing the good and, if she does lose the good, in rejoicing at this fact (passive malicious envy, *Schadenfreude*). The good may be of various kinds, such as personal qualities, social position, or material goods (one may add, *en passant*, that it is rather puzzling that human beings seldom envy other people their moral qualities, the *only* qualities which are really enviable). It seems to be most destructive in the case of personal qualities (it is then called 'existential'; cf. de la Mora 1987: 69), as here it is most difficult to overcome. It should be stressed that the object of envy need not always be the good *itself*; it may also be happiness or satisfaction that the good elicits in the rival. Furthermore, the malicious envy may take two slightly different forms, depending on whether the envier *above all* wants to achieve the good the other person (the rival) possesses (and only secondarily – if she cannot achieve this good – wants the rival to lose this good), or whether she *above all* wants to 'level down' the envied person, to see her deprived of the envied good, and only secondarily wants to achieve this good. It seems that malicious envy is especially wrongful in that second case. It may be illuminating to connect the above considerations about malicious envy with the analysis of human desire proposed by the French historian and anthropologist René Girard. He has argued that the structure of human desire always has three elements: it consists of the subject (of the desire), the object, and the mediator (rival), that is, the person imitated by the subject (Girard

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

1961). Accordingly, our desires always have a mimetic character: we desire something because the mediator (who may also be imagined, even supernatural) – whom we admire and wish to imitate – desires it. Thus, in every instance of envy, as Girard claims, there occurs an envier's fascination for the rival/mediator. It is a 'romantic lie' (*mensonge romantique*), so Girard's argument goes, that our desires can be autonomous, authentic, genuine, spontaneous, truly 'ours'. In reality we do not choose what we really want but what the other wants (Girard applies this analysis also to the emotion of love, claiming that love always has a mediator: we love a given person because someone else – the mediator/rival – loves this person as well). One may therefore say that on Girard's view (though Girard does not put it this way), behind our desire for an object there stands an existential envy directed towards the mediator, that is, a deeper (second-order) desire to be like the mediator, to 'take over' his or her being. It is therefore a mistake to claim that the object of envy precedes the rival: it is the rival (mediator), towards whom we feel existential envy, who precedes our desire for a concrete object. Girard seems to maintain that each (first-order) desire has the mimetic character, that is, that existential envy is the omnipresent and in fact the only form of envy. This is a radical and rather counterintuitive claim. But, assuredly, many people predominantly have the kind of desires – that is, mimetic ones – which Girard so insightfully examined.

Just envy contains two elements: (1) an unpleasant emotion (distress, pain, nuisance, and so on) felt by the subject at the thought that she does not possess the good and the rival does; (2) a desire that the rival lose the good – a desire which appears because, in the envier's view, it is unjust that the rival possesses this good. Due to component (2), just envy can be considered as a primitive form of a sense of justice. This kind of envy has positive effects: it may lead to engaging in fair competition; it may sensitize an agent to various manifestations of injustice. The distinction between just and malicious envy can be traced back to Hippias of Elis, who 'distinguished two types of envy, the just one, or the envy toward undeserving people when they receive honours, and the unjust one of those who envy good people' (de la Mora 1987: 4).

Benign envy is simply an unpleasant emotion (distress, pain, nuisance, and so on) felt by the subject at the thought that she does not possess the good and the rival does, and unaccompanied by any kind of desire that the rival lose this good. As D'Arms (2008) points out, benign envy is difficult to distinguish from a positive desire for a good, that is, from greed, or from admiration for the rival. It often motivates an agent to aspire to achieve, by just means (in a process of fair competition), the envied good.

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

Different types of envy seem to have different sources. Luis de Granada, the sixteenth-century Spanish mystic, argued in his book *Guía de pecadores* that 'the root of envy is pride, for pride cannot suffer the superior nor the equal' (quoted in de la Mora 1987: 52). This seems to be fully apt with regard to malicious envy. The malicious envier often seems to be concerned not so much with the good the envied person possesses, but rather with gaining superiority over the rival: his main problem is often his inability to bear the fact that his situation is inferior to that of the rival, not the mere lack of the good that the rival possesses. This desire for superiority may often flow ultimately from the feeling of existential inferiority, which the envier wants to hide from himself and others. Apart from pride and an inferiority complex, one could surmise that malicious envy may also flow from 'the will to power', or from unhappiness (which makes it hard to bear the happiness or the imagined happiness of the other person), or, as claimed by the great twentieth-century Spanish philosopher Miguel de Unamuno in his book *La envidia hispánica*, from 'spiritual idleness … mental shallowness and the lack of great intimate projects' (quoted in de la Mora 1987: 56). These sources may also play a part in the two other types of envy, but assuredly assume a weaker form there.

It is worth noticing that envy is most acutely felt towards people who are in some way similar to us – in whose place we could find ourselves; who are in our proximity. However, one can imagine a kind of *acute* malicious envy, which is felt even if 'there is no relation of similarity between the envied and the envier' (de la Mora 1987: 54). Another acute form of malicious envy is *ressentiment* – a combination of malicious envy and a sense of one's own impotence – which leads, as was perspicaciously described by Scheler (who was developing Nietzsche's insights from his *Zur Genealogie der Moral*), to the distortion of the values in question: the person who feels *ressentiment* will be inclined to undermine the values which she, due to weakness, cannot achieve. The result will be an inversion of values. Paradoxically, this may cause the disappearance of envy, for 'how can one be more happy if he possesses a counter-value? He must be a poor soul, worthy of compassion and scorn' (de la Mora 1987: 69).

Finally, let me raise the question of whether envy could have been preserved by natural selection. It may be the case that some forms of envy are the product of natural selection and some others are not. It is most difficult to justify the adaptive character of malicious envy, which is one of the greatest pathologies of the human spirit. It is not only ignoble and vicious to feel (which, of course, would not by itself imply that its existence is not probable in the light of evolutionary theory), but also

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

does not seem to bring any evolutionary advantage, as it is a highly self-destructive emotion. It decreases the envier's subjective welfare in a twofold manner: (1) malicious envy is an unpleasant vice – the envious person derives no satisfaction from envying the other person (the feeling of unpleasantness is a component of all forms of envy but it is incomparably stronger in malicious envy than in the two other forms of envy); (2) since it is radically antisocial, destroying all sympathetic connections between human beings, nobody wants to confess to possessing this vice – it generates shame and contempt. This reluctance to confess to malicious envy may also result from the fact that malicious envy is often the result of an inferiority complex. Confessing to envy would therefore amount to confessing that one feels inferior, that one's spirit is shrunken, mean, petty. This is the reason why malicious enviers so often engage in self-deception, trying to convince themselves that the emotion they feel is not malicious envy but, for instance, just envy or moral indignation. Malicious envy may also motivate the envier to take actions that will turn out to his disadvantage.

The above remarks are intended to show that it is highly improbable that malicious envy, given its self-destructive aspects, could have been preserved by natural selection. But they assume that malicious envy, as a pathology of human spirit, is quite frequent. The other argument for the nonadaptive character of malicious envy goes in a different direction: it assumes that malicious envy does not exist or exists much more rarely than is commonly supposed, and that what we consider as malicious envy usually turns out to be, on closer inspection, some other psychological phenomenon, such as just envy or moral indignation. Thus, our commonsensical picture (one shared by some philosophers, such as Arthur Schopenhauer) of human beings as frequently moved by malicious envy may be too dark; as Lars Svendsen put it, 'I trust that *Schadenfreude* is usually motivated by the feeling that suffering has been earned because the person did something evil' (Svendsen 2011: 104). Human beings may therefore exhibit a tendency to rejoice at other people's justified suffering, but not to rejoice at their unjustified suffering. This does not mean, of course, that there do not exist people who feel malicious envy, but they seem to be a negligible minority: 'hardly anyone', Svendsen claims, 'would enjoy seeing an innocent person executed' (Svendsen 2011: 104). This low frequency of occurrence of malicious envy would therefore be another argument in favour of the claim that it is not a tendency shaped by natural selection. To summarize – malicious envy is not a biological adaptation because: (*first line of argumentation*) even though it is quite frequent, it is too self-destructive to have been preserved by natural selection (it is therefore a pathology of human spirit, though a *frequent*

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

pathology); (*second line of argumentation*) it is a very rare phenomenon (so the explanation why it could not have been preserved by natural selection is, so to say, redundant). It seems that the types of envy that are part of our nature are just envy and benign envy. Unlike malicious envy, these two types of envy seem to bring evolutionary advantages. The evolutionary benefits of benign envy are obvious: it provides a powerful motive for an agent to pursue his cherished goals. Just envy, in turn, is a manifestation of the subject's unwillingness to accept injustice and thereby a signal to the other members of a society that the agent will not accept, for instance, distributions that fail to award her a reasonable part of a good being divided. Thus, just envy strengthens the subject's motivation to pursue the good she desires and constitutes a protection against others' attempts to take advantage of her.

2.3 Aggression

One can distinguish two general types of theory of aggression. One type assumes that aggression is not a product of biological evolution but of cultural or environmental factors. This type need not detain us further, as it is based on the entirely implausible denial of the role of natural selection in shaping our psychological/behavioural propensities. The other type assumes that our propensity for aggression constitutes a product of biological evolution, and thereby is 'built into' the human genotype. There are two basic variants of this view, which can be called 'pessimistic' and '(moderately) optimistic'. The pessimistic variant assumes that aggression is an innate instinct whose energy 'accumulates' and which has to manifest itself irrespective of whether such manifestation serves any adaptive goals or is simply destructive. This view was defended by Sigmund Freud (1995) and Konrad Lorenz (1996), for example. This theory of aggression is sometimes called 'hydraulic' because it implies that aggression is like water incessantly gathering in a tank: if its pressure reaches a certain level, it will overflow. There are two important implications of this account: first, that aggression does not have only a reactive, instrumental character – it may also be nonreactive, noninstrumental, spontaneous; second, that by claiming that aggression may be purposeless, it admits the existence of absolute evil, that is, of pursuing evil not as a means to an end but as an end in itself. Which of these two variants is supported by evolutionary theory? Arguably, aggression was for our ancestors an indispensable mechanism for survival and reproduction, which enabled them to solve concrete adaptive problems in their environments (such as defending one's own life and the lives of one's family against attack, protecting property, striving for social status,

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

discouraging sexual partners from infidelity). It was therefore reactive and instrumental rather than spontaneous and noninstrumental. Furthermore, it was not a single instinct but rather a group of distinct mechanisms which were preserved by natural selection to promote our genes. One can therefore say that the fact we have a propensity for aggression, that is, that we are not 'born pacifists', supports in some way the thesis that we are not genuinely moral, and the fact that our aggression is reactive supports in some way the thesis that we are not immoral. This account fits the picture of human nature as doubly ambivalent. Of course, there are some people who manifest pathological aggression – serving no practical purpose, directed against people who did them no wrong – but they are not numerous: they cannot be taken as typical exemplars of human nature.

One more point needs to be made. Irrespective of which theory of aggression we accept, we should be careful not to treat aggression as a key to understanding and explaining human evil. It is seldom a free-standing motive of evildoing; in point of fact, it can be such a motive only if it is understood in accordance with a noninstrumental – 'hydraulic' – theory of aggression (which is the least plausible of all theories of aggression). It is usually a part of a larger motivational structure (including, for example, various negative emotions). Furthermore, a motivational structure of evildoing does not have to contain aggression. Thus, it is hard to disagree with Mary Midgley that 'there are plenty of other ways of going wrong besides the aggressive way. To speak of all injustice as aggression seems to be a distortion of words caused by a mistaken attempt to narrow the problems of evil' (Midgley 1997: 75).

2.4 Self-Deception

The evolutionary theory teaches us that the human mind has not been designed to seek the truth as an end in itself, but only in so far as it helps to solve adaptive problems. As Michael T. Ghiselin put it:

> We are anything but a mechanism set up to perceive the truth for its own sake. Rather, we have evolved a nervous system that acts in the interests of our gonads … If fools are more prolific than wise men, then to that degree folly will be favored by selection. (Ghiselin 1974: 1)

This statement is excessively radical and provocative, because it may suggest that the human mind is not capable of freeing itself from the fetters of evolutionary interests. But it is plausible to maintain that

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

disinterested respect for the truth is not our natural tendency; we have to counteract our natural proclivities to reach this lofty attitude. Since the human mind is not a disinterested truthseeker, one can expect that it will be prone to generate various illusions, if this is the best way to solve a concrete adaptive problem. For instance, a well-established fact is that we are prone to self-deception, that is, to unconsciously concealing unpalatable information from ourselves and thereby distorting our picture of reality. As Michele K. Surbey put it:

> The self-deceptive mechanisms have potentially arisen as the outcome of a number of selective forces: natural selection for cognitive efficiency, maintenance of a positive outlook and hope in adversity, concealment of threatening thoughts that could limit adaptive responding, the ability to form mutually beneficial social alliances; kin selection for smooth family relationships; and sexual selection for success in mate selection and retention … At moderate levels, self-deception promotes mental health and effective coping with environmental demands. At extremely low or high levels, individual or collective self-deception may have detrimental consequences. (Surbey 2004: 140)

Thus, self-deception, if it is at a moderate level, brings advantages, which is why it could have been preserved by natural selection. However, its dominant functions are morally negative. These functions may be offensive or defensive. As to the offensive ones, self-deception facilitates deception: an agent can more easily deceive other persons if he can deceive himself as to his real intentions; if he believes them to be morally pure, and thereby shows no signs of anxiety, his deception becomes harder to detect. As was noticed by Mark Twain in the first volume of his *Autobiography*, 'when a person cannot deceive himself, the chances are against his being able to deceive other people'. If this is true, then Robert Trivers may be right in saying that

> until shown otherwise, we should assume that the intellectually gifted are often especially prone to deception and self-deception. Those who take pride in their alleged intellectual gifts or of their particular grasp might well contemplate whether they are also more regular liars and self-deceptors. They are expected to be better at it. (Trivers 2011: 91)

As for its defensive functions, since self-deception blocks the activity of an agent's conscience, it helps him preserve a positive image of himself in spite of evidence to the contrary. As such, it is an obstacle to moral self-improvement. One can distinguish at least four morally reprehensible manifestations of self-deception connected to serving the defence of positive moral self-image. First is pharisaism, that is, self-righteous

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

conviction about one's moral superiority over other persons, which may lead to rash moral judgements about other people – ascribing them evil intentions, interpreting their wrong behaviours as an effect of their supposedly evil character rather than of situational factors, and so on. Second is the so-called 'moralization gap', which is a partial, distorted perception of situations in which an agent is a victim or a perpetrator of an immoral act: in the former case an agent exaggerates the harm suffered; in the latter he belittles the harm done to another person. Third is hypocrisy, which may manifest in three different ways: in using different moral measures (respectively, more and less strict) in the evaluation of the other person's and one's own behaviour; in situations not clearly regulated by morality in choosing such actions which bring us benefits; and in situations clearly regulated by morality in justifying one's immoral action by saying that other people are morally even worse (the last two elements of hypocrisy were insightfully examined by Joseph Butler in *Sermon X* from his *Fifteen Sermons Preached at the Rolls Chapel*). Fourth is suspending or switching off our empathy: it is impossible to do harm to innocent persons without engaging in self-deceitful mental processes which make one believe, for example, that the innocent person is 'really' guilty, or that the person is not a 'full' human being but is instead more like an animal (animalistic dehumanization) or a machine (mechanistic dehumanization) (cf. Livingstone Smith 2007a, 2007b).

3. NEUTRAL TENDENCIES

I shall focus on three types of morally neutral tendencies, viz. (selected) emotions, hierarchical propensities, and a desire for self-transcendence. They are morally neutral in the sense that they can be put to bad or good (moral) use. But, arguably, they are biased in a morally wrong direction. Thus, if they are left in their 'natural condition', they are more likely to generate morally undesirable results than desirable ones.

3.1 (Selected) Emotions

With few exceptions (for example, empathy or envy), emotions are morally neutral in the sense that their moral evaluation, if at all possible, will usually depend on particular circumstances. They therefore cannot be evaluated *in abstracto* – that is, as types of emotions – as morally good or bad; if such an evaluation makes sense at all with regard to them, it would be perhaps apposite to call them 'morally neutral'. For instance,

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

grief, anger, disgust, or fear may be morally acceptable depending on their causes, their intensity, and the actions to which they lead. But even though these emotions are in themselves morally neutral, it seems that their natural condition is that of imperfection. Some examples will serve to illustrate this claim.

Excessive fear leads to cowardice, which may account for a large number of examples of evildoing. Anger, as was insightfully shown by Seneca in his treatise *De ira* (On Anger), easily gets out of control and begets reactions which are disproportionate to the cause (which is the sense of being harmed); in its excessive form it may be responsible not only for violent actions, but also for various primitive legal institutions (self-help or blood revenge) overcome only after hundreds of years of the evolution of legal systems. Disgust, in turn, is a biological adaptation which serves to protect the human body from harmful substances. This kind of disgust is called 'primary'. However, it can be easily extended to serve not only the protection of the human body's purity, but also the 'purity' of the human soul or person. This variety of disgust does not have to be unjustified: it seems perfectly proper with regard, for example, to sexual intercourse in public, or to incest. It is a kind of defence against treating human beings as having only an animal nature. But in the history of law and morality, such disgust tended to be extended in an entirely arbitrary way. For instance, one of the reasons why homosexuality or interracial marriage was penalized for such a long time was the psychological process that leads to (mistakenly) taking some people's 'gut reaction' of disgust to a certain type of act (for example, homosexuality or interracial sex) to be a signal of the moral impropriety of such an act. This is a special case of the process Paul Rozin (1999) called 'moralization' – the conversion of preferences into values. Moralization of disgust can be interpreted in three different ways (cf. Pizarro et al. 2011). On the first (strong) interpretation (assumed correct by Rozin himself), disgust is regarded as a moralizing emotion: 'morally neutral acts can enter the moral sphere by dint of their being perceived as disgusting' (Pizarro et al. 2011: 268); on the second, disgust is considered as a consequence of a moral judgement; on the third, it is an 'amplifier of a moral judgment', making morally wrong acts (that is, according to some independent moral rules) seem even more wrong. According to David Pizarro and his collaborators, the strong interpretation has weaker empirical support than the weaker ones. They also notice that:

> disgust cannot be sufficient for moralization to occur because there is a plethora of behaviors that are judged by most people as disgusting but not

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

> immoral, such as eating pig brains or picking one's nose in private. A credible defense of the claim that disgust exerts a moralizing influence would seem to require a plausible account of why it does not seem to moralize behaviors in most cases. One possibility is that disgust exerts a moralizing influence only on behaviors for which there already exist nonmoral proscriptive norms. In these cases, the pairing of disgust with (or the tendency to be disgusted by) the behavior might cause it to be 'pushed' into the moral domain … If this view is correct, one would expect moralization over time to occur only in the disgusting behaviors for which there are already conventional norms in place. (Pizarro et al. 2011: 128)

But even if one assumes one of the weaker interpretations, disgust can still be regarded as playing an important role in moral judgement – that of its entrenchment or amplification.

It should be stressed not only that the negative consequences of emotions may concern people other than the one experiencing the emotion, but also that the latter person may also suffer from the excesses of her emotions. Given this rather obvious fact, it seems surprising that in the philosophical literature devoted to personal welfare so little attention has been paid to emotions: none of the three dominant theories of welfare which provide criteria for answering the question of how well a person's life is going for that person stresses emotions as a factor relevant for generating welfare. The mentioned dominant theories are the following: *hedonic theory*, according to which wellbeing consists in the greatest balance of pleasure over pain; *desire theory*, according to which well-being consists in preference satisfaction; and *eudaimonic theory*, which provides an objective list of goods (for example, knowledge, friendship, pleasure, autonomy) supposedly necessary for wellbeing. As regards the relations between emotions and welfare, it seems that emotions, when left in their 'natural state', constitute a serious obstacle for experiencing welfare. At least three arguments speak for this claim. First, the natural state of our emotional setup seems to be disordered: we have many emotions which pull us in various directions, and thereby jeopardize our autonomy. Second, some emotions are an important source of welfare and some others are an obstacle for experiencing welfare. But even those emotions which are a source of welfare (such as love or joy) are easily transformed into emotions which reduce welfare (for example, into grief, hatred, disappointment). Third, evolutionary biology teaches us that emotions have evolved to facilitate the realization of specific tasks which have nothing to do with our happiness. Accordingly, from the standpoint of evolutionary biology, happiness is not an ultimate goal of human actions but a means to realizing a more fundamental evolutionary goal: survival and reproduction (and therefore spreading our genes). Humans

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

are therefore not 'destined' to be happy, but rather to effectively pass on their genes to subsequent generations. They are 'designed' to experience happiness only so long as it serves evolutionary goals. This conclusion – that emotions need to be subject to 'therapy' if they are to bring about a 'positive balance' of welfare – was reached independently by many ancient thinkers, such as Plato, the Stoics, the Sceptics, the Epicureans and the Buddhists. These considerations should not be interpreted as onesidedly critical towards emotions. Emotions are, let me repeat, neutral (with some exceptions): they can be put, not only to bad uses, but also to good ones. For instance, rage (anger) seems to be an indispensable part of our moral sensitivity; as Hannah Arendt noticed:

> Rage and the violence that sometimes – not always – goes with it belongs among the 'natural' human emotions, and to cure man of them would mean nothing less than to dehumanize or emasculate him … Absence of emotions neither causes not promotes rationality. Detachment and equanimity in view of unbearable tragedy can indeed be terrifying, namely, when they are not the result of control but an evident manifestation of incomprehension. In order to respond reasonably one must first of all be moved and the opposite of emotional is not rational, whatever that may mean, but either the inability to be moved, usually a pathological phenomenon, or sentimentality, which is a perversion of feeling. Rage and violence turn irrational only when they are directed against substitutes. (Arendt 1970: 64)

Nonetheless, if one wanted to formulate a general judgement about emotions, it seems there would be justification in saying that they seem to be biased in the wrong direction and, if left in their 'natural state', constitute an obstacle to reaching a high level of personal welfare.

3.2 Hierarchical Propensities

In his analysis of primitive egalitarianism, the evolutionist Christopher Boehm (2001) convincingly argued that human beings have been endowed by natural selection with hierarchical propensities: the will to dominate and an aversion to being dominated. In Chapter 3 I present his account of how these propensities may have given rise to the primitive egalitarian structures. Here, I would like to make some more general observations about these hierarchical propensities.

First, it seems that – apart from the will to dominate, that is, the desire for power (*libido dominandi*), and an aversion to being dominated – hierarchical propensities embrace also the propensity to obey authority (the dominant individual or the dominant group). The human tendency to obey authority, just like the two other hierarchical propensities, seems to

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

be a biological adaptation: it is highly probable that in ancestral environments it constituted an individual's fitness-enhancing disposition in situations of conflict with more dominant members of a group. If, in ancestral environments, a conflict between a relatively powerless individual and a relatively powerful agent arose, the former agent was faced with the choice between fighting with the latter (and thereby being killed or sustaining high costs of a battle) or deferring to it (and thereby recognizing its higher place in the social hierarchy, but simultaneously avoiding the costs of fighting). Clearly, in such a situation the adaptive strategy of the relatively powerless agent would be to defer to the relatively powerful agent. Even though the result of this strategy for the relatively powerless agent would be only the second-best one (the best result for him would be the unattainable domination over the relatively powerful agent), it would be better than the result of a fight with the relatively powerful agent. Accordingly, it is plausible to maintain that a deferential strategy which directs an agent to obey more dominant members of a group has evolved. This strategy is present in many species, for example in chimpanzees, which are especially sensitive to signals of social power. It is supported by such cognitive capacities as, for example, the capacity to determine the relative power of the members of a group, or the capacity to calculate the costs and benefits of getting into a conflict with a member of a group.

Second, the ethical status of the three hierarchical propensities may be different. There seems to be little controversy that aversion to being dominated and the tendency to obey authority are indeed morally neutral. The case of desire for power is much more disputable. One could plausibly argue that it should be regarded as an immoral tendency, related to *superbia* and *hybris* (and thereby giving rise to what I have called *hybris*-based egoism).

Third, it must be admitted that exact psychological relations between these three hierarchical propensities are far from clear. For instance, according to Hannah Arendt, the desire for power goes in tandem with the tendency to obey authority, and both are incompatible with the aversion to being dominated (desire for freedom). As she put it: 'The psychological truth is that the will to power and the will to submission are interconnected; a strong disinclination to obey is often accompanied by an equally strong disinclination to dominate and command' (Arendt 1970: 46). But this view is controversial. One may just as well argue that the desire for power is especially acute among those people who are especially averse to being dominated, and thereby have 'a strong disinclination to obey'.

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

Fourth, it should be emphasized that the claim that hierarchical propensities are part of our 'common' biological nature, that is, were shaped by natural selection, is perhaps more controversial than claims regarding other propensities interpreted in this book as a product of biological evolution. For instance, according to many scholars, not all people exhibit a tendency to obey authority, only those who have a specific – pathological – character structure, called 'sadomasochist' by Erich Fromm and 'authoritarian' by Theodor Adorno. Such people exhibit rigid adherence to conventional values, are submissive to authority figures and are contemptuous towards people lower down the social ladder. They also desire power, are fascinated by it, and admire the powerful. They tend to divide people into two broad categories: those who possess power and those who are deprived of it. Furthermore, they are aggressive towards outgroups and inimical to introspection, reflection, creativity. They exhibit a tendency towards superstition, stereotyping and projecting their own, negative, unconscious impulses onto the outer world. They also often exhibit exaggerated concern with sexuality. As was emphasized by Fromm, this kind of character structure is a symptom of a personality defect: a person with a strong self, capable of realizing her various potentialities, will not exhibit a desire for domination. It should be stressed, however, that Fromm's views need not be regarded as inconsistent with the claim that hierarchical propensities have been shaped by natural selection, and thereby are shared by most people. One may argue that people with the authoritarian – sadomasochistic – personality belong to the margins of the bell curve of the distribution of hierarchical propensities: while most possess these propensities to some degree, people with this personality type exhibit them much more.

3.3 Desire for Self-Transcendence

The distinction between self-assertive and self-transcending tendencies was introduced by the renowned writer and philosopher Arthur Koestler in his book *The Ghost in the Machine*. Self-assertive tendencies are connected with the competition within each species for territory, mates, dominance, food (Koestler 1975: 70). They therefore serve individual, egoistic interests. Self-transcending tendencies, which are present only in human beings, are integrative: they are aimed at making an individual a part of a larger whole, and manifest themselves in 'the longing to belong' (Koestler 1975: 243). It might seem that the role of self-transcending tendencies is positive, because, as Koestler wrote, 'while the self-assertive emotions (rage, fear, anger, jealousy) narrow the field of consciousness (passion is not blind but blinkered), the self-transcending

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

emotions expand it, until the self seems to dissolve in the "ocean feeling" of mystic contemplation or aesthetic enchantment' (Koestler 1975: 218). However, according to Koestler, self-transcending tendencies are in fact more dangerous than self-assertive ones. He claims that for most people, 'self-transcendence' is realized not in a creative way (for example, by finding an outlet in artistic or scientific activity), but by identification with a tribe, caste, nation, party, or church. It therefore takes the form of a 'regression to an infantile form of self-transcendence – immersion in the group mind is a kind of poor man's self-transcendence' (Koestler 1975: 240). Koestler made the bold (though assuredly exaggerated) claim that the main source of evil is self-transcending (integrative), not self-assertive, tendencies. What is therefore really dangerous, in his view, is the better part of human nature – its integrative/self-transcending tendencies: 'The crimes of violence committed for selfish, personal motives are historically insignificant compared to those committed *ad majorem gloriam Dei*, out of a self-sacrificing devotion to a flag, a leader, a religious faith, or a political constitution' (Koestler 1975: 234). In Koestler's view, the worst crimes of human beings 'have nothing to do with the seven deadly sins. The eighth sin, deadlier than all, self-transcendence through misplaced devotion, is not included in the list' (Koestler 1975: 236). This sin leads to the obliteration of individual egoism but also the creation of group egoism: anger, vengefulness, hate, fear do not disappear but take a different form: they are triggered by a concern not with individual interests but with group interests. The type of aggression they lead to is named by Koestler 'secondary' (as opposed to 'primary', which is caused by self-assertive tendencies). Thus, it can be said that 'the glory and the tragedy of the human condition both derive from our powers of self-transcendence' (Koestler 1975: 251).

Koestler's judgement of human beings is too harsh. For instance, he writes, in an apocalyptic tone:

> When one contemplates the streak of insanity running through human history, it appears highly probable that *homo sapiens* is a biological freak, the result of some remarkable mistake in the evolutionary process. The ancient doctrine of original sin, variants of which occur independently in the mythologies of diverse cultures, could be a reflection of man's awareness of his own inadequacy, of the intuitive hunch that something along the line of his ascent has gone wrong. (Koestler 1975: 266)

But his gloomy considerations have an undeniable value: they stress the often forgotten fact that the potential for evil inherent in our self-transcending tendencies is not to be neglected. It may be useful to put his account of the sources of evil in a broader context.

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

Arguably, the evil generated by self-transcending tendencies – by a desire to realize some moral ideal – is an instance of a more general category of what may be called 'moralistic evil', that is, evil justified (in the conscience of the perpetrator) by his conviction that what he is doing is morally good. In point of fact, many evil actions are regarded by their perpetrators as justified or even required by morality, especially by justice. James Gilligan's opinion is assuredly too strong but, certainly, rightly emphasizes the neglected aspect of the aetiology of crime: 'All violence is an attempt to achieve justice, or what the violent person perceives as justice … The attempt to achieve and maintain justice or to undo and prevent injustice, is the one and universal cause of violence' (Gilligan 1997: 11, 12). In similar vein: 'I have yet to see a serious act of violence that was not provoked by the experience of feeling shamed and humiliated, disrespected and ridiculed, and that did not represent the attempt to prevent or undo this "loss of face"' (Gilligan 1997: 110). The appeals to justice may therefore take place not only at the abstract level (of which Koestler writes) – that is, the level of fighting for the wellbeing of one's own group or some abstract ideal – but also at the level of individual relationships, when, for instance, a victim's kin frankly believe that justice requires them to make a bloody revenge, or when a perpetrator commits an evil act to overcome shame caused by lack of self-esteem as a result of being humiliated and treated as inferior, and not uncommonly rooted in abuse experienced as a child (cf. Gilligan 1997: 42–7). That moralistic evil is more widespread than is usually supposed seems to be confirmed by research showing that 'harmdoer guilt was higher following accidental as opposed to intentional transgression' (McGraw 1987: 247). The fact that premeditated evil elicits weaker feelings of guilt can be most plausibly interpreted by assuming that before the evildoer performed his action, he justified it morally before his conscience. Too much (superficially understood) morality may therefore have tragic consequences. According to Lars Svendsen, 'an overwhelming amount of the evil in the world happens because of love – love of self, love of family and friends, love of country, love of our abstract ideal, love of a leader' (Svendsen 2011: 124). Of course, one should be sceptical of general claims such as Gilligan's that 'all' violence is a desperate way to compensate for one's lack of self-esteem (and originally for a feeling of being humiliated or not being loved), or that the 'overwhelming amount' of evildoing is a result of excessive love. There are, of course, many other psychological sources of evildoing, such as *hybris* leading to a specific form of egoism, other forms of egoism, envy, and so on. But it cannot be doubted that moralistic evil, as Gilligan and

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

Svendsen rightly point out, is a much more common phenomenon than is usually supposed.

4. MORAL TENDENCIES: EVOLUTIONARY ETHICS

There is no doubt that the tendencies analysed in this section are positive from a moral point of view. However, the 'natural' (that is, shaped by natural selection) form they take is imperfect, diverging from the requirements of genuine ethics. What they give rise to is evolutionary ethics – a rudimentary, imperfect form of ethics, composed of four elements: empathy, kin altruism, tribalism, and reciprocal altruism. The term 'evolutionary ethics' is used here, as before, in a twofold sense: with reference to both the psychological, moral tendencies, and to moral rules which are 'correlates' of these tendencies.

4.1 Empathy

One should be sceptical as to the quality of our empathy in its 'primitive' form, that is, as a product of biological evolution and undeveloped by reflection and moral exercise. In order to justify this thought, I shall distinguish three types of empathy – the perfect, the truncated and the contaminated – and shall argue that the 'primitive' form of empathy is either truncated or contaminated. I shall start with a description of the type of empathy which is part of 'genuine ethics'.

Perfect empathy is the combination of three elements: cognitive empathy, affective empathy and the tendency to take empathic actions. Cognitive empathy, in its full-fledged form, is the capacity for a comprehensive and ethically proper understanding of other people's emotions. It should be stressed that cognitive empathy may, in particular cases, be confined to understanding the reasons for the other person's emotions, not the emotions themselves, since the reasons for a given emotion may fail to generate the emotion itself. To give an example: a person who has suffered the loss of reason 'laughs and sings perhaps, and is altogether insensible of his misery' (Smith 2007: 13), but the proper reaction to the loss of reason would be sorrow. An important fact about cognitive empathy is that it becomes ethically valuable only when it is coupled with affective empathy. Affective empathy, in its full-fledged form, is the tendency to emotionally respond to other people's emotions *in an ethically proper way*, which means that it implies cognitive empathy. Cognitive empathy is sometimes called 'empathy' *tout court*, while affective empathy is called 'sympathy'; the former is 'feeling

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

someone's pain' (or pleasure), the latter 'feeling *for* someone who is in pain' or rejoicing with someone who is joyous (cf. Slote 2007: 13). The above – full-fledged – forms of cognitive and affective empathy should be distinguished from their simple forms, meaning only, respectively, an understanding of the other person's emotions, and responding to these emotions with similar emotions (sorrow to sorrow, joy to joy). The empathic action tendency is the tendency to undertake an ethically proper action (for example, helping or consoling) as a result of the emotional response to another person's emotions. The phrase 'ethically proper', which occurs in the above definitions of the constituent parts of perfect empathy, implies that *perfect empathy is not a freestanding ethical capacity*. An understanding of the other person's emotion is only partial, and may be distorted if it does not include knowledge of the causes of the emotion. For instance, the sorrow of the agent who experiences this emotion because of the loss of the beloved person is essentially different from the sorrow of an agent who experiences it because of the success of his or her enemy. One can therefore speak about a true understanding of the other person's emotion only if one knows its causes and, additionally, is capable of distinguishing between the justifying causes (that is, the reasons) and the nonjustifying causes of a given emotion. As was mentioned earlier, understanding may be confined only to the reasons for a (potential) emotion of the person with whom we empathize if that person, being unaware of these reasons, does not experience the emotion itself. Affective empathy, in turn, will be ethically proper if it means responding with sorrow to another person's justified sorrow, with joy to another person's justified joy, and with indifference or sorrow to another person's unjustified sorrow or unjustified joy (the sorrow of the empathic person would not be caused in this case by the other person's emotion as such, but rather by the fact that the other person reacts in a morally inadequate way). One should also add that affective empathy need not be a response to an emotion: it may be a response only to the other person's reasons for a (potential) emotion – reasons of which she may be unaware (for instance, we empathize with a person who has lost a close relative but is not yet aware of that loss and so does not yet feel sorrow). It is clear that perfect empathy is not necessary for moral action: one can easily imagine a nonempathic agent strictly following moral rules and thereby undertaking moral actions. But it must be admitted that perfect empathy provides an especially strong motivation to undertake moral actions. Furthermore, if one were to provide an agent-based, not action-based, moral evaluation of agents undertaking moral actions, then the evaluation of a perfectly empathic agent would undoubtedly be higher than that of an unwavering but nonempathic follower of moral rules. One

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

may notice in passing that it follows from the above analysis that the ethics of *perfect* empathy might arguably be a plausible ethics, but it would not be a pure ethics of empathy, because the very concept of perfect empathy, as defined in this chapter, is a normative one – that is, it implies a set of moral rules prescribing when, in what circumstances, and towards whom empathy is demanded; what actions should be taken as a result of the empathic response to the other person's feelings; and how to decide conflicts of empathic impulses. That an ethics of empathy cannot be built was the thesis held also by Max Scheler (1980: 234–53), who stressed that for empathy (or sympathy, the term he used) to have ethical value, it must be empathy with emotions *deserving* empathy, and in order to distinguish the emotions deserving empathy from those undeserving of it we must have prior (independent of empathy itself) conceptions of justice, blame, responsibility.

Truncated empathy lacks one of the free elements of perfect empathy, or contains one or more of them but only in a truncated form. Truncated empathy may therefore take various forms, ranging from those which are morally reprehensible to those which are morally commendable. The morally reprehensible form of truncated empathy is the combination of cognitive empathy with the absence of affective empathy, that is, the combination of a capacity to recognize the other person's emotions with an incapacity to react to them. This kind of combination is characteristic for psychopaths, though, arguably, is not sufficient for creating a psychopathic personality. As Simon Baron-Cohen (2011: 154) noted, the psychopath, apart from displaying this combination, must additionally be 'morally negative'. This implies that the combination of cognitive empathy with the absence of affective empathy is, by itself, not sufficient for generating a psychopathic personality; for this personality to arise, some kind of profound indifference to moral rules or to other people's wellbeing is needed. This indifference to other persons can be described in terms proposed by Martin Buber (2012): as the tendency to treat other persons as things, or to replace the personal relation based on '*You*' with the impersonal relation based on '*It*' (of course, this is still not an explanation of this tendency, but simply highlights its different aspect). The morally commendable form of truncated empathy consists of cognitive empathy and affective empathy in their simple forms (that is, not supported by moral rules). Clearly, this combination is desirable: a person who displays these two forms of empathy exhibits sensitivity to the other person's feelings, and this sensitivity is valuable in itself. But, as mentioned above, this kind of sensitivity, if unsupported by the knowledge of moral rules and the motivation to analyse the causes of the other person's emotions, is likely to lead to morally improper actions if

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

the other person's emotions are unjustified (if, for example, the person empathized with feels sorrow for immoral reasons).

Contaminated empathy arises by admixing one of the following four amoral or nonmoral elements to perfect empathy, viz. (1) *a feeling of relief*, that is, thankfulness at the contrast between our good fortune and the sufferer's misfortune; (2) *a feeling of anxiety* about our own good fortune arising at the sight of the sufferer's misfortune; (3) *a feeling of superiority* over the other person (*pity*); and (4) *personal distress*: the unpleasant feeling that arises at the sight of the sufferer's sorrow, caused by the image of suffering rather than by anxiety about one's own good fortune.

It is to be emphasized that all four motives contaminating perfect empathy are self-regarding – they have as their ultimate end the well-being of the agent, not the interests of the other person. They can also function as freestanding motives, in which case they can be regarded as various forms of what may be called *pseudoempathy.* The first two of these motives were analysed by Samuel Butler in *Fifteen Sermons Preached at the Rolls Chapel*, in his famous polemics using Hobbes's definition of pity as 'fear felt for oneself at the sight of another's distress'. This was nicely summarized by Charlie Dunbar Broad in the following passage:

> He [Butler] points out (a) that, on this definition, a sympathetic man is *ipso facto* a man who is nervous about his own safety, and the more sympathetic he is the more cowardly he will be. This is obviously contrary to fact. (b) We admire people for being sympathetic to distress; we have not the least tendency to admire them for being nervously anxious about their own safety. If Hobbes were right admiration for sympathy would involve admiration for timidity. (c) Hobbes mentions the fact that we tend specially to sympathise with the troubles of our friends, and he tries to account for it. But, on Hobbes's definition, this would mean that we feel particularly nervous for ourselves when we see a friend in distress. Now, in the first place, it may be doubted whether we do feel any more nervous for ourselves when we see a friend in distress than when we see a stranger in the same situation. On the other hand, it is quite certain that we do feel more sympathy for the distress of a friend than for that of a stranger. Hence it is impossible that sympathy can be what Hobbes says that it is. Butler himself holds that when we see a man in distress our state of mind may be a mixture of three states. One is genuine sympathy, i.e., a direct impulse to relieve his pain. Another is thankfulness at the contrast between our good fortune and his ill luck. A third is the feeling of anxiety about our own future described by Hobbes. These three may be present in varying proportions, and some of them may be wholly absent in a particular case. But it is only the first that any plain man means by 'sympathy' or 'pity'. (Broad 1930, p. 53)

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

At the end of this passage Broad treats 'sympathy' (or compassion) and 'pity' as identical. But there are good reasons to treat them as distinct. The distinction was neatly made by Philip Mercer (1972: 18) and recalled by David E. Cartwright (1988). Compassion, corresponding to what I have called 'perfect empathy', is an altruistic – *other-regarding* – virtue having as its ultimate aim the wellbeing of the other person. Pity, by contrast, is egoistic, or *self-regarding*: it is expressive of contempt towards the other person and aims at heightening one's self-esteem and feeling of power. It embraces cognitive empathy but it lacks affective empathy: there is no feeling of sorrow at the other person's suffering or, at best, the feeling of sorrow is mixed with joy at the possibility of increasing one's self-esteem and the feeling of one's own power. Cartwright characterizes it in the following way:

> the pitier is superior in status to the pitied. We do not pity those we respect or those we judge superior to ourselves – unless we wish to level them by devaluing their status … By pitying them, I elevate myself. I boost my feelings of self-esteem by exercising my pity. The same is true when I pity someone who is suffering … The sufferer is helped, but helped in order to enhance my feelings of superiority. In these regards, pity is self-regarding. If we have general duties to respect others, pity incites their violations. If the moral goodness of beneficence is due to a desire to pursue another's well-being, the help rendered out of pity is not morally good. (Cartwright 1988: 559)

Pity is not morally good because it does not lead to 'acts of beneficence for the right sorts of reasons' (Cartwright 1988: 560).

Let me summarize the main conceptual distinctions introduced above:

(1) *Perfect empathy* consists of ethically grounded cognitive empathy, ethically grounded affective empathy, and the tendency to undertake morally adequate actions as a result of affective empathy.
(2) *Truncated empathy* appears if one of the elements of perfect empathy is lacking, or if at least one of its elements is not supported by the clear knowledge of moral rules (that is, it is not ethically grounded, or a tendency to undertake actions is not guided by the clear knowledge of moral rules).
(3) *Contaminated empathy* is perfect empathy combined with one of the four self-regarding motives: a feeling of relief that another, not myself, is the victim of suffering; anxiety about my own future; a feeling of superiority (pity); personal distress.
(4) *Pseudoempathy* occurs if any of the four 'contaminating' motives appears as a freestanding motive.

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

As mentioned above, it was the view of a number of philosophers, including Hobbes and Nietzsche, that perfect empathy does not exist; that in each empathic action one can always seek out some egoistic motive. But this view does not seem convincing: common experience, the famous experiments on the empathy–altruism hypothesis carried out by C.D. Batson (1991), and the strong philosophical arguments raised by Butler support the view that perfect empathy (characterizing a genuinely ethical person) does exist (though certainly is not a frequent phenomenon). Furthermore, as Cartwright aptly remarks (in the spirit of Butler's critique of psychological egoism defended by Hobbes),

> simply showing that an agent may derive pleasure, relieve feelings of sorrow or grief, feel self-satisfied, or better about oneself for helping another, is not to show that the action is self-regarding. In the same regard, simply arguing that in one sense all interests are mine, in the sense that I possess them, does not show that this is a self-regarding interest. What he had to show was that the end of the action was the agent's pleasure, feelings of self-esteem or superiority. (Cartwright 1988: 564)

Thus, Nietzsche could have not proved that the agent can never have the other person's wellbeing as his end. What he was able to prove was that in many cases of purported altruistic empathy, a different – egoistic – motivation may in fact be operating. But one must concede that there is some reason in Nietzsche's claim: it seems that perfect empathy rarely or never appears in an uncontaminated form. Let me repeat, however, that Nietzsche believed that perfect empathy never exists, even in a contaminated form; he believed that all forms of empathy are at bottom egoistic, that is, are forms of pseudoempathy. This is an extreme and implausible view. The plausible view, however, is that our empathy, in the form shaped by biological evolution, is *very far from being perfect*. In addition to the fact that it is imperfect, very often contaminated by various forms of pseudoempathy, it has three other defects.

First, its 'natural', that is, evolutionary-shaped level of intensity, is not very high. To put it more precisely, this feature (intensity of empathy), as most other features, has a bell-curve distribution over the entire population, and the level of empathy which is exhibited by most people does not seem to be very high. The evolutionary reason for this fact is that people with a very high level of empathy would be unable to take sufficient care of their own interests and people with a very low level of empathy would pose a threat to a group and be eliminated; as a result, neither very high nor very low-level empathy could become dominant in the population. La Rochefoucauld was right in saying that 'nous avons tous assez de force pour supporter les maux d'autrui'.

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

Second, primitive empathy is partial, that is, sensitive to factors which are often ethically irrelevant. We are more likely to empathize with those who are similar to us (in various regards – for instance, in terms of physical appearance, social status, race), who are more cute, who are our kin, or who are simply more visible. Especially interesting manifestations of the last of these four types of partiality are the 'identifiable victim' effect and the 'collapse of compassion' effect. The former consists in greater willingness to help a concrete, individualized victim than a 'statistical' victim (cf. Jenni, Loewenstein 1997). The latter (which can be interpreted as a partial explanation of the former) consists in decreased empathy as the number of victims increases (Slovic 2007). Such a decrease may take one of two forms – either less or more radical. In the less radical form, empathy is assumed to be subject to the law which is analogous to the law of decreasing marginal utility: successive victims elicit smaller 'marginal' increases of empathy (or, alternatively, the value – utility – assigned to each successive human life is smaller). This means that the 'total empathy' (that is, empathy for the entire group of victims) increases disproportionately slowly with an increasing number of victims. In the more radical form, an increasing number of victims causes a decrease in 'total empathy'. It is hard to say which of the two versions of the effect really occurs. Slovic thinks it is the more radical. But even if it is the less so, it still illustrates the limitations of our empathy, or, as Slovic puts it more strongly and aptly, 'a fundamental deficiency in our humanity' (Slovic 2007: 79). The limitation is in fact an illustration of the priority we give to our own interest over the interest of others: our small loss moves us much more strongly than other people's great loss. As Adam Smith memorably wrote:

> Let us suppose that the great empire of China, with all its myriads of inhabitants, was suddenly swallowed up by an earthquake, and let us consider how a man of humanity in Europe, who had no sort of connexion with that part of the world, would be affected upon receiving intelligence of this dreadful calamity. He would, I imagine, first of all, express very strongly his sorrow for the misfortune of that unhappy people, he would make many melancholy reflections upon the precariousness of human life, and the vanity of all the labours of man, which could be annihilated in a moment. He would too, perhaps, if he was a man of speculation, enter into many reasonings concerning the effects which this disaster might produce upon the commerce of Europe, and the trade and business of the world in general. And when all his fine philosophy was over, when all these humane sentiments had been fairly expressed, he would pursue his business or his pleasure, take his repose or his diversion, with the same ease of tranquility, as if no such accident had happened. The most frivolous disaster which could befall himself would occasion a more real disturbance. If he was to lose his little finger tomorrow,

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

> he would not sleep tonight; but provided he never saw them, he will snore with the most profound security over the ruin of a hundred millions of his brethren, and the destruction of that immense multitude seems plainly an object less interesting to him, than this paltry misfortune of his own. (Smith 2007: 136–7)

It is important not to interpret these facts onesidedly, treating them as illustrating the darkness of the human heart. One can present them in a less pessimistic light. For instance, one may argue that the phenomenon of 'the collapse of compassion' simply shows that our empathy 'evolved to protect individuals and their small family and community groups from great, visible, immediate danger; this effective system did not evolve to help us respond to distant, mass murder' (Slovic 2007: 87). Or one may say it is a rational response of human agents observing a tragedy of a large group: they 'expect the needs of large groups to be potentially overwhelming, and, as a result, they engage in emotion regulation to prevent themselves from experiencing overwhelming levels of emotion' (Cameron, Payne 2011: 14). However, even granting that our defective empathy may be understandable and perhaps even justified, we can still maintain that this kind of empathy is far from being ethically proper.

Third, our 'natural' empathy is unreliable – it can be rather easily suspended or weakened. Without going into an analysis of various ways in which the perpetrators of great atrocities 'switched off' their empathic reaction (for example, by dehumanizing the victims), let me mention less dramatic examples from our everyday experience. Most of us know all too well that the strength of our empathy depends on such morally irrelevant factors as whether we are tired or in a rush; we also know how easily empathy can be overshadowed by strong emotions (not only negative, such as anger, envy, hatred, or a desire for revenge; positive emotions of happiness may also engender a high level of self-centredness). As was mentioned in Section 2.4 of this chapter, suspension of empathy does not seem possible without the mechanisms of self-deception.

Having reached the end of my analysis of empathy, I would like to make two additional observations.

The first concerns the definitional relations between evil and empathy. Simon Baron-Cohen (2011) defined evil as an erosion of empathy. This definition is inadequate for two reasons. Since it is focused on the internal state of the perpetrator of evil, it does not make precise what concrete 'effects' in the external world must be caused by an action for it to be called 'evil'. Arguably, in order to make a proper account of evil, we must combine the account of the internal state (motivation) of the

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

agent with the account of the action this motivation leads to. But even as an account of the evil motivation, this definition is not apt – it is too narrow. It is based on the assumption that only empathy can block our evil actions. But this is implausible. An unempathic agent may nonetheless be virtuous or obey rules. Thus, if one wanted to enumerate various possible 'brakes' on immoral behaviour, one should mention at least two additional 'mechanisms' apart from empathy, viz. virtues and rules.

The second observation concerns relations between empathy and feelings of guilt. As it turns out, they are quite strictly connected to each other. The strength of our feelings of guilt depends not only on the type of our wrongdoing (as it should) but also on the type of person whom we wronged. The feelings are stronger if we wronged a person with whom we want to keep close and positive relations; one may say that we may have sincere feelings of guilt only towards those persons with regard to whom we can be truly empathic. Consequently, the more empathic persons are, the more likely are they to be guilt-prone. It is also worth recalling Dostoyevsky's observation in his *Memoirs from the House of the Dead* that he met no criminals with feelings of guilt. This is a sadly predictable reaction: if an agent decided to commit an evil act, and thereby considered it the right thing to do, he should not be expected to feel guilty because of this act (unless he changed its moral evaluation, which occurs rarely). Guilt may also take other improper (false) forms (cf. Tournier 1962: 64–98). It may be insincere: an agent may only pretend to feel guilty. Or an agent may sincerely feel guilty, but his guilt may be irrational. The irrationality is extreme if the agent feels guilty about what, on virtually all moral theories, he cannot be blamed for: for example, for being healthy (when others are ill), for having survived a natural catastrophe, for having survived a concentration camp. It is less extreme, though still irrational, if an agent failed to perform a supererogatory action, that is, he did not decide to expose himself to a high risk in order to help a person in danger. Arguably, this broad or overly sensitive variety of the feeling of guilt is not a product of natural selection; rather, it arises as a result of specific personal experiences. But what may have been such a product is the ease with which guilt is suspended or 'silenced' – various techniques of neutralization of conscience, with which we come up spontaneously and with little effort (for instance those described by Gresham Sykes and David Matza (1957), viz. the condemnation of the condemners, the appeal to higher loyalty, the denial of injury, responsibility, or the victim) prove to be distressingly effective.

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

4.2. Kin Altruism

The theory of inclusive fitness, originally put forward by William D. Hamilton (1964) as a correction of the theory of natural selection, is contemporarily regarded as an integral part of the theory of natural selection. This theory states that an individual's reproductive success embraces not only her/his personal reproductive success (as was implied originally by the theory of natural selection), that is, the number of her/his offspring, but also the reproductive success of other individuals discounted by their coefficients of relatedness to the individual. The theory of inclusive fitness explains why individuals display altruistic behaviour towards their kin, that is, it partly resolves the so-called 'problem of altruism' (the existence of altruistic behaviour was a puzzle for evolutionary biologists, including Darwin himself, before the theory of inclusive fitness was propounded). To put it more precisely: the theory of inclusive fitness explains the existence of biological kin altruism, as it leads to the prediction that natural selection will favour the psychological mechanisms which make people behave altruistically towards their kin – because they bear a high percentage of their genes – and that the degree of altruism will be in inverse proportion to the degree of genetic relatedness. However, kin altruism also has a dark side. It leads to nepotism and favouritism towards kin. It may also be argued that its consequence (or, rather, a side effect) is tribalism (of which I write more in the next section). Kin altruism also seems to lie at the basis of the mechanism called 'discriminative parental solicitude' (cf. Daly, Wilson 1999: 37). This mechanism consists in a greater likelihood of adult persons delivering their finite resources to those children who are most capable of turning them into reproductive success. The existence of this mechanism accounts for the fact that children unrelated to an adult person and children related to that person but who are physically or mentally handicapped are at increased risk of abuse on the part of that person. The empirical data confirm the hypothesis that the psychological mechanisms responsible for child abuse have biological foundations. For instance, they reveal one of the consequences of 'discriminative parental solicitude', which is called 'the Cinderella effect': this is that 'the probability that a stepfather or boyfriend of the mother will kill an unweaned infant is nearly *one hundred times* greater than is the probability of death at the hands of an infant's genetic father' (Jones, Goldsmith 2005: 453). These data remain telling even if we allow for the obvious fact that the overwhelming majority of stepfathers do not kill their stepchildren and that most step-parents also manifest true concern for the children. This is arguably

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

one of the most dramatic examples of 'secondary evil' generated by one of our moral tendencies, viz. kin altruism.

4.3 Tribalism (Group Altruism)

Tribalism consists in treating one's group as better than other groups, and consequently holding negative emotions towards the other group. It is typical of a primitive communal mentality and is aimed at ensuring the survival and prosperity of the group. It is based on the assumption that the members of one's own group and the members of other groups are different in essential aspects, which justifies undertaking sacrificial actions for the sake of the former and manifesting deep hostility towards the latter. The distinct features of tribal ethics are therefore exclusivity – the circle of people who deserve moral respect is narrowed down to the members of one's own group – and (misguided) idealism – people are expected to make great sacrifices for the sake of their own group. Darwin described these ethics in the following way:

> Actions are regarded by savages, and were probably so regarded by primeval man, as good or bad, solely as they obviously affect the welfare of the tribe – not that of the species, nor that of an individual member of the tribe. This conclusion agrees well with the belief that the so-called moral sense is aboriginally derived from the social instincts, for both relate at first exclusively to the community. The chief causes of the low morality of savages, as judged by our standard, are, firstly, the confinement of sympathy to the same tribe. Secondly, powers of reasoning insufficient to recognizing the bearing of many virtues, especially of the self-regarding virtues, on the general welfare of the tribe. Savages, for instance, fail to trace the multiplied evils consequent on a want of temperance, chastity, etc. And, thirdly, weak power of self-command; for this power has not been strengthened through long-continued, perhaps inherited habit, instruction, and religion. (Darwin 1871: 80)

There is controversy among evolutionists as to the origins of tribal thinking. There are two main arguments for the claim that tribalism is embedded in our biological predispositions, that is, that there exist biological predispositions whose correlates are the rules that constitute tribal ethics (for example, such rules as 'treat your own group as better than the other groups', 'be ready to sacrifice your own individual good for the sake of the good of the group to which you belong'; 'do not trust the members of other groups'). The first argument appeals to the theory of kin altruism, the second to the theory of group altruism. These two evolutionary explanations of the emergence of predispositions underlying tribal ethics need not be viewed as mutually exclusive: they can be

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

regarded as complementary. Let me present these arguments in somewhat greater detail.

The first argument says that the emergence of predispositions underlying tribal ethics can be explained by the theory of kin altruism. According to this argument, therefore, tribalism would in fact be a result of an extension of the circle of the objects of ethical concern, from a very narrow circle of the closest kin to a still narrow, though slightly broader circle of the members of one's group. The extension may be interpreted in terms of mistake or side effect: we treat members of our groups as our kin because the mechanisms of distinguishing between kin and non-kin are imprecise and fallible.

The second argument states that these predispositions are the product of the mechanism of genetic group selection. The theory of genetic group selection asserts that group selection may favour the emergence of the propensity to sacrifice oneself for the sake of the group, because groups consisting of individuals endowed with this propensity are likely to fare better than and prevail over groups consisting of individuals without such propensity. This explanation assumes, then, that the propensity to undertake sacrificial acts for the sake of the group (let us call this propensity 'group altruism') may have evolved, because fitness losses incurred by individuals endowed with such a propensity are compensated by the superior performance of the group to which they belong. It is worth mentioning that some fragments from Charles Darwin's writings suggest that he was inclined to accept the idea of group selection as an explanation for the evolution of human faculties or psychological features; see, for example, his statement that 'A tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to aid another, and to sacrifice themselves for the common good, would be victorious over most other tribes' (Darwin 1871: 89). However, this explanation is controversial. The basic argument advanced against it is that if an egoist appears within a group of group-altruists, he will have a higher chance of survival and reproduction than group-altruists, so that after a few generations the initially altruistic group will become dominated by egoists. This is the so-called 'subversion from within argument' against the idea of group selection, raised by proponents of the view that the basic unit of selection is gene, not group (cf. Williams 1966; Dawkins 1976). It should be noted, however, that in recent times the idea of genetic group selection has been rehabilitated to some extent. As Elliott Sober and David S. Wilson (1998) have demonstrated, should some conditions be met (and, as they argue, the satisfaction of these conditions is facilitated by cultural evolution), it is possible that between-group

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

selection for group altruism (that is, for the propensity to sacrifice oneself for the group) will prevail over within-group selection for egoism (that is, for the propensity to exploit group-altruists). The conditions are as follows: genetic variation between groups must be greater than genetic variation within groups; the speed of between-group selection must be relatively higher than the speed of within-group selection; in order to hinder the process of within-group selection for egoism, the groups must grow, disperse, fission, and integrate in optimal times. The problem with the idea of genetic group selection is that these conditions are difficult to satisfy. For instance, the processes that uphold genetic variation between the groups and select between the groups are usually weaker than the processes that disrupt genetic variation and select within groups (for example, the processes of migrations among groups).

It is worth noticing, at the end of this section, that the question of whether human nature is primarily individualistic or social (communal) – whether, in the early stages of history, human beings had felt themselves primarily to be separate entities with their own interests, or parts of a larger whole, of a community (with their individual ego having been, so to say, not distinct) – is closely related to the question (still not answered clearly by evolutionary biologists) about the *depth* of the rootedness of tribalism in human nature. But it seems it can be reasonably assumed that tribalism is not a superficial part of human nature (and is closely related to the self-transcending tendencies discussed in Section 3.3).

4.4. Reciprocal Altruism

A precise characterization of the human tendency to engage in reciprocal altruism was provided by game theorists, who have identified the kind of behaviour (strategy) which is most effective in generating beneficial (utility-maximizing) outcomes in relations of social exchange, and thereby is likely to have been preserved by natural selection. In their view, the bulk of human beings are neither universal cooperators – 'suckers' – nor universal defectors – 'cheaters' – but rather tit-for-tat (*TFT*) players (cf. Axelrod 1984). *TFT* is strikingly simple, as it starts with a cooperative move and then imitates an opponent's last move. This strategy is therefore nice, reciprocal, (moderately) forgiving, not (maliciously) envious, and clear – that is, someone using this strategy is never first to defect, punishes defectors and therefore cannot be exploited by them, forgives after an opponent's single period of cooperation, and does not want to gain more than her opponent. Being easily recognizable, this strategy is effective in generating cooperation. The insight that, in a series of Prisoner's Dilemmas (modelling the relations of social exchange), *TFT*

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

is the strategy that fares best and thereby has the highest chance of reproductive success in reciprocal relationships is a very intuitive one, for it accords with the commonsense observation that repeated interactions create the possibility to punish defectors and thereby discourage them from defecting.

Two additional remarks on this strategy are in order. First, *TFT* can be seen as embodying moral rules of a specific character, viz. conditional and prudential. These rules can be presented in two equivalent ways. One can say that *TFT* embodies the rule 'Do unto others as you would have them do unto you *only if* others do unto you as they would have you do unto them'; or that it embraces the following rules: (1) start by cooperating; (2) do good to those who do good to you and do harm to those who do harm to you; (3) be (moderately) forgiving; (4) do not be (maliciously) envious (cf. Singer 1995: 129–53). Thus, since *TFT* does not prescribe unconditional moral action, it cannot be regarded as embodying a moral rule *sensu stricto*; it is a prudential rule, that is, one aimed at serving our own long-term interests. Second, an interesting problem connected with the research on the *TFT* strategy concerns the role of emotions in shaping the behaviour of agents involved in the Prisoner's Dilemma-like interactions. One way in which emotions may affect the agent's behaviour is by supporting the mechanism of reciprocal altruism: certain emotions (for example, anger, guilt, gratitude, or a 'second-level' emotion – empathy) may mimic the behaviour of the *TFT* players. Furthermore, emotions can affect the agent's utility function in such a way that it changes the game in which the agent was initially involved. For instance, the game that was initially the Prisoner's Dilemma may become (say, under the influence of the emotion of empathy, which makes the agent derive disutility from her opponent's misfortune) a game in which a rational choice is cooperation rather than defection; thus, emotions may induce agents to act cooperatively even in the one-shot Prisoner's Dilemma (by transforming this game into a different one). Thus, the relations of social exchange may have created a strong selective pressure for the appearance of certain emotions: these emotions are supposed to 'move' the agents to take actions they would have taken were they able to carry out the rather complicated calculations that a fully rational player is expected to make. They therefore serve as proxies for the *TFT* strategy: agents who act on these emotions act *as if* they played the *TFT* strategy (cf. Trivers 1971).

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

5. EVOLUTIONARY ETHICS AND THE NECESSITY TO OVERCOME IT

Evolutionary ethics embraces ethics of empathy, family (kin altruism) ethics, tribal ethics, and reciprocal altruism ethics. This ethics, 'encoded' in our biological nature, is primitive: its level of impartiality is low, and as a result it may generate morally undesirable (from the standpoint of genuine ethics) consequences. More generally, the ambivalence of the elements of evolutionary ethics can take two forms: first, a lack of moral purity; second, undesirable side effects. Kin altruism may lead to nepotism, favouritism of kin, and discriminatory parental solicitude (and the consequent 'Cinderella effect'). Tribal ethics is based on the assumption that members of one's own group and members of other groups are different in essential aspects, which justifies undertaking moral actions for the sake of the former and manifesting hostility towards the latter (it leads therefore to ingroup cooperation and outgroup competition). It cannot be regarded as a part of genuine ethics because of its exclusivity (a narrow definition of the circle of people who deserve moral respect), and thereby its arbitrariness, which leads to a double standard in moral action and moral evaluation. Reciprocal altruism (tit-for-tat) ethics cannot be regarded as embodying moral rules *sensu stricto:* its prescription of cooperation is conditional and prudential, that is, aimed at serving the agent's long-term interests, and is therefore contaminated by egoism. The ethics of empathy is also imperfect, since empathy in its natural form is fragile (easy to 'suspend'), limited (partial), often blind to morally relevant factors, and impure, that is, contaminated by various forms of pseudoempathy (the feeling of relief, the feeling of anxiety, the feeling of superiority over the other person, or personal distress). Accordingly, in order to reach full moral development, one must pass from evolutionary ethics to genuine ethics. At the level of genuine ethics one therefore abstains not only from primary evil (which is absent already on the level of evolutionary ethics) but also from secondary evil, that is, evil caused by moral (as well as neutral) tendencies shaped by natural selection.

In the considerations pursued in this chapter I have implicitly assumed that human beings are endowed with freedom of will, that is, the capacity to choose different actions in a given moment. Thus, an agent who has chosen an action at a given moment has free will if she could have chosen some other action at that moment. The capacity for freedom resides in or is strictly connected with our capacity for symbolic thinking and anticipating the consequences of our actions,

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

which, though itself the product of natural selection, enables us, so to say, to transcend the boundaries of biology – to liberate ourselves from the force of our evolved propensities (cf. Ayala 2010). As a result, humans are not 'closed' in their individual perceptions (which can be called 'the first-level representations of the world'). Owing to their capacity to use abstract notions, they create symbolic representations of the world (which can be called 'second-level representations of the world'). This capacity (arguably unique among animals) creates, as one may call it, an ontological crevice between humans and the world, and thereby fundamentally detaches the two from each other. This is the anthropocentric (strongly humanistic) assumption that cannot be conclusively proved (and which is not popular in the contemporary climate of opinion: cf., for example, Pietrzykowski 2016). I shall confine myself to invoking three arguments in its favour. The first one says that 'our personal experience indicates that the possibility to choose between alternatives is genuine rather than only apparent' (Ayala 2010: 324); the second that 'when we confront a given situation that requires action on our part, we are able mentally to explore alternative courses of action, thereby extending the field within which we can exercise our free will' (Ayala 2010: 324). Third and furthermore, as was forcefully argued by Kant, it seems that one cannot speak meaningfully about responsibility for actions, and about (moral) evil as such, if we do not assume that human beings are free. Similar thoughts have been defended by many other philosophers. For instance, Lars Svendsen claimed that 'moral evil begins with free will, while the causes of natural evil are purely natural. Biology doesn't have the capacity to define the moral concept of evil and at most can only explain away the phenomenon' (Svendsen 2011: 21). Indeed, one cannot biologically 'define' the concept of moral evil, because biology cannot help us distinguish between good and evil and because it cannot explain the ultimate source of evil (free will). It must be added, however, that it can be immensely helpful, even indispensable, in explaining our 'gravity', that is, the psychological propensities which 'push' us towards (nonbiologically defined) evil actions (which indeed could not be called 'evil' if not finally approved or selected by our free will).

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

6. INSTEAD OF A SUMMARY: ON SIMILARITIES BETWEEN THE EVOLUTIONARY ACCOUNT OF EVIL AND THE CHRISTIAN AND KANTIAN ACCOUNTS

The choice of the Christian and Kantian accounts of evil as a background against which to see more clearly the doubly ambivalent view of human nature is not accidental: there are deep similarities between the evolutionary, Christian and Kantian accounts.

6.1 The Christian Account

Since Christianity gave rise to many different currents of thought, one cannot maintain that there exists one specifically Christian account of evil. I shall present two such accounts (which are dominant in the Catholic variant of Christianity): a less and a more optimistic. They share the following two assumptions: that evil has a negative character, and that evil flows from freedom. I shall have more to say about the second assumption in the Epilogue to this book. Here I shall focus on the first one, which is a basis of division between these two accounts. The thesis about the negative character of evil may be understood morally or ontologically (these two senses are to a large extent separate from each other). From the standpoint of the comparison of the evolutionary and the Christian accounts of evil, the moral sense is essential. I shall therefore make this my focus, and make only some general comments on the ontological sense.

The moral sense is expressed by the statement that our will, as St Augustine put it in *De Civitate Dei*, is not an efficient but a deficient cause of evil acts: it causes evil not because it chooses evil as such, but because it chooses an inferior good. Evil is therefore negative in the sense that it consists in *not choosing* a superior good and in choosing an *inferior* good. This claim implies that human will is, basically, directed towards goodness; as the Thomist philosopher Josef Pieper put it,

> every sin consists in longing for a passing good: in every sin there lurks the desire to have and to enjoy, which is in fact why sin can never become infinitely bad, a pure evil. Because the reality of the world is good, no human deed can ever take on the character of being definitive evil, guilt, or sin. (Pieper 2001: 57)

It is noteworthy that two interpretations of this statement are possible: a less and a more optimistic one. On the less optimistic interpretation, this

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

statement is simply another way of saying that human beings do not commit absolute evil; that choosing evil is always choosing an inferior good. If understood in this way, it is, of course, fully compatible with the view of human nature as doubly ambivalent. One may, however, interpret it more optimistically (as the Thomist thinkers tend to do), viz.: that each of the natural human appetites, in which evil is ultimately rooted, are either moral (basically good) or morally neutral, and that thereby there do not exist natural tendencies which could be called 'immoral'. This is a controversial and *too* optimistic claim, since, as I have argued, human beings were endowed by nature with immoral tendencies. So there is full consistency between the less optimistic variant and the doubly ambivalent view of human nature, and partial consistency between the more optimistic variant and the doubly ambivalent view. However, even in the latter case, essential consistency exists, because the doubly ambivalent view assumes that none of the immoral tendencies are of such a kind as to deserve to be called 'absolutely evil'. What is more, there is also a distinct note of pessimism even in the more optimistic variant of the Christian account. Even though it asserts that our natural appetites are good, and our will directed towards goodness, it also claims (like the less optimistic variant) that the present situation of man (due to his contamination by original sin) is imperfect: it is that of, to use the Christian term, *cor curvatum in seipsum*, wherein man's heart becomes turned towards itself, he makes himself his final end, self-love is his dominant motive, and he may even commit *peccatum mortale*, or 'turning away from God' (I shall return to this problem in the Epilogue). Thus, the Christian and the evolutionary account of evil share essential points: that human nature is flawed, but not fundamentally flawed (because there does not exist absolute evil), and that as such it must be (morally) transformed.

To have a more complete picture of the Christian account of evil, two additional remarks are in order.

First, there is also a different, teleological, understanding of the moral sense of the Christian claim that *malum est privatio*, that is, that evil is negative. If we assume that human beings are supposed to realize a certain God-given purpose (*telos*) and thereby fulfil their morally understood nature, then, should an agent act in a wrongful way, he will fail to reach what he ought to reach (his *telos* as a human being). As a result of his wrongdoing, his nature suffers a *deficiency*, a *lack*; his (empirical) nature is, so to say, incomplete, since it did not live up to its task. But this understanding is harder to reconcile with the scientific approach to morality, as it invokes the notion of *telos* – typical for Aristotelian metaphysics – that is alien to contemporary science.

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

Second, the ontological sense of thesis about the negative character of evil is that whatever exists is good. This claim may be justified in two different ways. It is a direct consequence of the Christian claim that what was created by God is good. It may also be viewed as a consequence of the claim that whatever exists has measure, form, order – things that are unquestionably good – and thereby itself is good (things create a hierarchy of goodness: they may be more or less good, but nothing is bad in itself). Accordingly, once all goodness (and thereby measure, form and order) is removed, nothing 'positive' is retained – nature itself disappears. Both lines of argumentation can be discerned, for instance, in St Augustine's *De natura boni*. Clearly, these theological considerations go beyond the perspective assumed in this book, but they are, of course, not incompatible with it: there is no inconsistency in being at the same time an evolutionist and a theist.

6.2 The Kantian Account

Immanuel Kant did not use the term 'double ambivalence' in his account of human nature, especially regarding the sources of human evil, but his account is similar to the one proposed in this book in two main points: first, it is neither optimistic nor pessimistic, but ambivalent, and implies that human beings have certain propensities for evil but not for absolute evil, that is, they do not pursue evil for its own sake; second, it states that a moral transformation is necessary for human beings to become truly moral.

In his book *Die Religion innerhalb der Grenzen der bloßen Vernunft*, Kant distinguishes three types of human propensity towards evil. The least serious type is frailty (*fragilitas*), which is an agent's weakness in observing the rules he has accepted; it is a weakness of will consisting in the fact that even though an agent wants to comply with the rules *for the right reasons* (that is, simply because they are moral rules), he fails to realize his intention because of the greater motivational force of some other (amoral or immoral) motives. Frailty is therefore yielding to a temptation to follow one's amoral or immoral inclinations, and thereby to violate a moral rule. A more serious type is impurity (*impuritas*, *improbitas*): an agent is 'impure' if she observes moral rules not only 'out of duty', that is, just because they are moral rules, but also for some other reason; for instance, she may be telling the truth not only because it is a moral duty but also because she is afraid that her lying will be detected and punished. Impurity is therefore a tendency to contaminate the motives of our moral actions by including among them immoral or amoral motives. As a result of this contamination, our moral actions are

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

only legal (in accordance with the duty), not moral in the strict sense (done only out of duty). The improper reason for action may in fact be stronger that the proper one – that of acting for the sake of duty. However, Kant does not analyse more closely the relation between 'proper' and 'improper' reasons in the motivation set of a morally 'impure' agent (that is, he does not analyse the problem of the gradation of impurity); for instance, he does not examine whether proper and improper reasons operate simultaneously, or whether the stronger reason 'switches off' the weaker one. It is noteworthy that, according to Kant, impurity is morally worse than frailty even though an impure person acts in accordance with a moral rule and a frail person fails to do so. This 'ranking' of propensities for evil can be explained by the fact that the frail person tries to perform the moral duty for the right reason (even if she fails in her efforts), while an impure person performs the moral duty for the (partly) wrong reason, and, for Kant, the crucial factor for the moral evaluation of an act is its motivation. In order for this ranking to be plausible, however, frailty cannot consist in yielding to *any* temptation to act amorally or immorally; it must be a temptation to do something which is only moderately wrong, such as to experience sensual pleasure. Kant does not make this point clear, but only in this way can his ranking be defended. The most reprehensible type of propensity for evil is wickedness or perversity (*vitositas*, *corruptio*), which consists in an agent prioritizing egoism (self-interest, self-love) over the moral law, and following the moral law only if it serves his own interest. It is therefore a tendency to invert the proper hierarchy of motives of action (there does not exist any reason, according to Kant, which would justify deviation from the moral law). Kant does not analyse at greater length the relations between impurity and wickedness, which may be interpreted in two different ways. On one interpretation, the difference between them consists in that, in the case of conflict between self-interest and moral law, the impure agent gives priority to the moral law, whereas the wicked agent gives priority to self-interest. In other words, in the case of such a conflict, the moral law plays no role in the motivational set of the wicked person, and plays a decisive role in that of the impure person. On this interpretation, one cannot be at the same time impure and wicked. But this interpretation implies that an 'impure' agent becomes 'pure' (that is, 'silences' his egoism) in the case of a conflict of motives, which is contradictory to the assumption that his motives are always contaminated. Besides, it would be rather strange that an agent who cannot act for the right reason when his motives point, so to say, in the same direction, can act for the right reason if they point at different courses of action. The second – more plausible – interpretation of the relations between

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

impurity and wickedness assumes that impurity implies wickedness: the very fact that the agent is impure, that is, acts in accordance with the moral duty only if he has some additional motive (other than the willingness to obey the duty) to do so, implies that in the case of the conflict between moral duty and self-interest she will give priority to the latter. Impurity and wickedness would therefore be two sides of the same coin: the term 'impurity' stresses the fact that the agent can never follow moral duty for the right reasons (he always needs an additional – nonmoral – motive to act morally), and the term 'wickedness' stresses the fact that in the case of the conflict between moral and nonmoral motives, the agent will give priority to the latter. The latter fact seems to follow from the former because a motive which can never act independently (one such motive is willingness to follow a moral rule) cannot have priority over the motive which can act independently (one such motive is self-interest). By way of digression, it may be interesting to invoke Lars Svendsen's apt observation that Kant's account of propensity for evil, though extremely insightful, does not embrace two other – all too frequent – propensities for evil, viz. propensity for moralistic (idealistic) evil, in the case of which 'the agent believes that he represents the good and his victims are evil', and propensity for thoughtless (stupid) evil, in the case of which 'an agent does not stop to think about whether his actions are good or bad' (Svendsen 2011: 137).

In his analysis of human propensities for evil, Kant used the term 'radical evil'. The relation between this term and the three propensities for evil is not entirely clear. It seems that radical evil in a narrow sense is wickedness, and in a broader sense is also weakness and impurity. Irrespective of which of these two is assumed, radical evil in the Kantian sense proves to be a relatively moderate form of evil: 'radical' does not mean 'extreme' or 'absolute'. Kant does not maintain that the reason why human beings infringe upon moral duties is that they derive some demonic pleasure from the very fact of infringing upon moral duty; they are not therefore 'absolutely evil', that is, they do not commit evil just for the sake of committing evil. Kant straightforwardly denies that human beings may take the violation of moral rule to be the reason for their actions. The question arises why Kant, who denies 'absolute evil', employs the apparently misleading term 'radical evil' (resembling 'absolute evil') to describe human nature. Taken etymologically, however, the term is by no means misleading: as Kant reminds us, 'radix' means in Latin 'root', and the most typical human propensity for evil – impurity – consists precisely in contaminating the 'root' of our moral actions, viz. their motivation. Furthermore, the claim that the propensity towards evil is 'radical' means also that it is deeply rooted in human nature. It is not

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

clear, however, whether it can be removed. Kant seems to have believed this possible, given that he believed in the *moral necessity* of a moral transformation of an agent. In Kant's philosophy, that something is morally required implies that it is morally possible; the principle 'Ought implies Can' is taken by him literally, as licensing the derivation of a factual statement (of what people can do) from a normative statement (of what they ought to do), or as licensing the transition from a normative statement (being a point of departure) to a factual statement. This is a nonstandard usage of this principle. In its standard usage, the principle is treated as a way of evaluating the plausibility of an 'Ought' statement by comparing it with a factual statement (being a point of departure) saying what people can do. In this usage, Kant's reasoning from 'Ought' to 'Can' is usually treated as unjustifiable – just like the reverse derivation, from the factual statement to the normative one. Kant would only agree that there is no transition from 'Is' (and its special case: 'Can') to 'Ought'.

Let me summarize. Kant's account of human evil is neither optimistic nor pessimistic. It is not optimistic because Kant admits that there exists in human nature a tendency to evildoing. But it is not pessimistic either, because Kant clearly states that human evil is not absolute evil: the human will is good, that is, it never wants to commit evil just because it is evil, and the worst evil which it can commit (prioritizing self-interest over the moral law) is far from being extreme. Human nature is therefore ambivalent, and as such needs to be 'transformed' if it is to become truly moral. Individual human beings must undergo a moral transformation/revolution, which will enable them to follow moral rules for proper reasons, that is, simply because they are moral rules. This moral revolution can be described, on the grounds of Kant's moral philosophy, in various ways: as coming to have a pure moral motivation (that is, performing the moral law just because it is the moral law); as gaining moral character, that is, a deep unity of personality, through a deep internalization of moral rules; or as developing a tendency to recognize the infinite value of each human agent, and thereby to treat him always as an end in itself and never merely as a means. In Kant's view, human beings are in a position to undergo such a transformation because they have free will. Kant also maintains that the very existence of various types of human propensity to evil is a result of a free decision by each individual human being (and that thereby we are responsible for having it), but he never explains how this is possible; he admits that this is a mystery which can only be made (partly) understandable if we invoke the theological doctrine of original sin. This is a somewhat puzzling fragment of Kant's considerations which I shall not analyse. I shall limit

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access

myself to noticing that the responsibility for our various types of propensity to evil would be plausible if it were understood not as responsibility for their arising but for their preservation: if we can remove our propensities (as Kant seems to believe), then we can be held responsible for them.

Wojciech Załuski - 9781786436504
Downloaded from Elgar Online at 01/27/2022 11:54:28PM
via free access