EVOLVING AUTONOMY:

THE MUTUAL SELECTION OF SOCIAL VALUES

INTRODUCTION

With the publication of John Rawls' *A Theory of Justice*,[1] the social contract model of political justification was resurrected and revised, in the hope that it would lend rational coherence to the modern liberal conception of justice based on fairness in the distribution of social benefits. If it could be shown that agents would be rationally motivated to opt for a society governed by co-operative, egalitarian principles, rather than risk their personal welfare in an unequal competition with their neighbours, then the legitimacy of the liberal state would be established on firm foundations. However, insofar as this legitimacy is reached only by divesting individual agents of any insight into their mutual and competitive interests, powers or personal histories, it has no motivational force for actual human beings with minds evolved to regulate thought and behaviour via inherited categories, heuristics and norms. Therefore, a truly viable social contract must be informed by empirical knowledge of these evolved conditions, to avoid adopting moral and political principles that are inherently impractical ('ought' implies 'can') or unsympathetic to the scope of human beings' physical and emotional sensibilities.

In *Morals by Agreement*,[2] David Gauthier lifted the 'veil of ignorance' that undermined the success of Rawls' project, enabling agents to reach agreement by bargaining on the basis of their known dispositions. Nevertheless, Gauthier's agents still proceed by ideal principles of rationality that ignore the cognitive and motivational constraints of practical reasoning. It is only when these *pragmatic* elements are factored into the bargaining situation that contracting agents are then able to determine the most rational procedures for selecting *theoretical* principles of justice that are best adapted to meeting fundamental human needs and interests. However, before such principles and procedures can be agreed upon by autonomous and rational human agents, it is first necessary to provide an empirical account of the actual conditions which allow these capacities to continually evolve and function to meet those basic human needs and interests. Accordingly, Part One comprises two chapters which together provide both the empirical evidence and theoretical foundations for the evolutionary approach that will later be used to show the *adaptive* value of principles and procedures endorsed under this new contractarian model.

Given the evident flexibility required to accommodate the evolution of human needs and interests, Chapter One is concerned with identifying the two evolutionary conditions necessary for the continuing maintenance of human agency i.e. the functional capacities of (a) autonomy and (b) inductive reasoning. Following Dennett[3] and Waller[4] autonomy is defined by the organism's capacity for self-control which, although itself the product of natural selection, is crucial to further enhancing the organism's scope for exploring new opportunities to modify its responses to the environment, or to render ecological conditions more conducive to its mode of life. A number of prominent biologists,

behaviourists and ethologists[5] have emphasised the organism's learning capacity as a selectional force in determining the direction of its development. Most important in this organic selection process is the co-evolution of autonomy and inductive means/ends reasoning, as both capacities work to synergistic effect on development. This is shown by some examples of the increasingly adaptive behaviour in different species' phenotypes.

Having seen how the evolution of these adaptive learning capacities is vital to the well-being of individual agents, Chapter Two shows how these same powers have been socially adapted for mutual benefit in the evolution of moral and political systems. This is done by examining recent evidence from historical and anthropological research and is further supported by findings drawn from game-theoretic models. Revealing the evolution of reciprocal altruism as the most evolutionary stable strategy (ESS) for individuals in all surviving societies, such models are presented as strong evidence for considering the merits of a social contract in terms of its pragmatic and evolving function.

The biological and anthropological research of Part One is put forward as the background for commencing the more overtly philosophical discussion of Part Two, whose unifying theme is to show how such evolutionary factors support both a meta-ethical and normative theory of the inherent pragmatism of moral agency, as an indispensable feature of any viable moral and political system. On this basis, Chapter Three defends a meta-ethical analysis of the inherently teleological and prescriptive function of moral judgements and the indispensable normative value of reciprocity. In the process, alternative moral theories are examined in the light of the evolutionary account, invalidating those that are fundamentally non-consequentialist, while showing neo-Aristotelian, agent-relative theories to be most adaptable to the evolved structure of moral thinking.

Evolutionary accounts of normative ethics are inevitably confronted by the perennial problem of fact/value distinctions and the related naturalistic 'fallacy'. Chapter Four examines and affirms the formal validity of the logical is/ought distinction. However, using Dewey's model of reflective means/ends reasoning to reveal the pragmatic unity of thought and action, the *logical* distinction is shown to be an irrelevant impediment to the evolving *psychological* process of evaluating possibilities for both individual action and social interaction. Developing Dewey's theory of evolutionary pragmatism to emphasise the role of organic selection noted in Chapter One, an holistic account of the mutual selection of facts and values is put forward to demonstrate the continuity of the evolutionary process to incorporate biological facts and cultural values. In this context, Baldwin's and Piaget's evolutionary theories of social and moral development are reexamined in the light of current research in developmental and social psychology, which shows remarkable support for Dewey's approach.

This discussion of the agent's role in the selection process leads to the central question of determining the essential conditions of agency from the standpoint of the individual agent. As grounds for generating moral rights to those goods necessary for maintaining the status of a 'prospective, purposive agent' (PPA), in *Reason and Morality*[6] Alan Gewirth identifies freedom and well-being as essential for the pursuit of any purpose, irrespective of its content. In the context of the preceding evolutionary analysis of agency as a vital developmental process, Gewirth's argument and his 'Principle of Generic Consistency' can be seen to provide a logical basis for justifying individual rights to such common agency needs. To that extent, Chapter Five assesses and endorses the validity of his

central argument, while drawing attention to problems with his generic conception of agents' needs, in the light of the organic selection process described in the previous chapter. As that process has shown that moral values are necessarily discovered and adopted in a particular and evolving social context, every prospective, purposive agent must rationally evaluate their agency needs from within that inherited context, rather than from the generic perspective described by Gewirth.

As the inheritance of established facts and normative value systems both enables and constrains the agent's opportunities for developing their own individual capacity for moral reflection and action, the continuing *evolutionary* conditions of agency must have an objective value for any prospective, purposive agent. This is the conclusion reached in Chapter Five by modifying Gewirth's generic argument to instead identify both autonomy (rather than freedom *simpliciter*) and a level of well-being conditional to the *particular* agent's continuing interests as necessary goods for their prospective purposes. This particularist conditionality in warranting individual rights to agency goods might seem to undermine any rational basis for contractual agreement on mutual rights. On the contrary, from an ontogenetic and pragmatic perspective, social and moral reasoning is constrained within certain strategic and dynamic parameters which can serve to set limits to the legitimacy of mutual rights claims. Chapter Five and Part Two concludes by illustrating these strategic and dynamic dimensions of rational and moral choice with contrasting examples.

Parts One and Two are primarily concerned with establishing a conception of agents and their needs as developing in accordance with evolutionary processes of organic selection – a process which emphasizes the role of agents' own choices and actions in both adapting the social environment to meet their needs and adapting their needs to the demands of the environment. For this reason, the political and moral principles that such agents would mutually select to govern their behaviour must take a long-term, strategic and dynamic perspective of how those principles would best protect and preserve their autonomy and well-being. In remodelling the social contract to reflect these organic effects upon their continuing agency needs, principles selected should ideally reflect the present and future autonomy safeguarded *by* putting them into practice, as well as the actual autonomy expressed *in* the rational choice of those principles. Thus, the chapters in Part Three are directly concerned with setting out the decision theories and epistemic principles best suited to ensuring that the evolving parameters of an organic social contract will genuinely reflect 'the rational choice of autonomy and the autonomy of rational choice'.

With a view to proposing a contractual basis for determining common rights to such *adaptive* agency needs, Chapter Six presents the principles and methods of rational coherence which can ideally serve as grounds for extending the legitimacy and scope of Gewirth's 'Principle of Generic Consistency'. As illustrated by Thagard's 'multicoherence' theory, moral and political principles and norms can be quite rationally chosen in terms of their inductive coherence, thereby avoiding the problem of building an entire theory of generic moral rights on a single deductive principle such as Gewirth's. To that end, Hare's account of universalisability is also defended as setting logical constraints on consistency in moral reasoning. However, it is only the evolutionary perspective which reveals the pragmatic constraints and 'affordances' which affect the coherence of choices available to actual agents in the real world, as opposed to the abstraction of reasoning behind a 'veil of ignorance'.

Here, it is Reed's version of ecological psychology with its concept of 'affordances' (resources in the material or social environment) which gives full weight to the pragmatic coherence afforded by adopting different values, rules and practices in an evolving social environment. Ultimately, all of these rational coherence factors are proposed with a view to generating a level of 'reflective equilibrium' that is sufficiently wide to set global conditions for reaching agreement on the basic moral and political principles required by agents concerned to protect their evolving needs. Therefore, the chapter concludes by outlining a pragmatic model of reflective equilibrium that is specifically designed to reflect the organic, social dimensions inherent in the rationality of individual choices.

As a mechanical process favouring the development of behaviours which enhance the organism's survival prospects, natural selection alone does not make any moral distinctions. Gewirth cites this 'problem of specificity' in evolving the capacity for autonomy to produce harmful or unethical behaviours, as well as those that are ethically agreeable. In Chapter Seven, this objection is explicitly dealt with by invoking the principles of ecological and dynamic coherence that function as dialectical conditions in the cultural evolution of moral systems. The dynamic perspective of agency[7] and the significance of ecological affordances[8] will often combine to specify a choice of means and ends whose moral benefits are already well established. Nevertheless, the experimental indeterminacy in the ontogenetic and cultural evolution of moral principles and beliefs may seem to equally permit the adoption of immoral or seriously maladaptive behaviours. To the contrary, this chapter will show that it is only by affirming the *adaptive value of human agency* that such behaviours can best be *avoided*. By acknowledging and anticipating the material and psychological organic selection pressures which induce such harmful, dysfunctional behaviours, it becomes possible for individual agents to initiate preventative or remedial actions against such effects.

The political and moral benefits that result from the preservation and development of such adaptive agency conditions can therefore serve as the rationale for a social contract along the lines of Gauthier's, but where agents' practical recognition of the universality of evolved needs leads them to develop dispositions toward mutual aid that extend beyond the bounds of local relationships[9] toward a cosmopolitan system of ethics directed at the growth of agents' capacities to meet those needs.[10] On this basis, Chapter Eight first outlines the process whereby such agents' select the most rationally adaptive method of evaluating competing moral theories or principles, as measured by their conformity with the principles and norms of rational and strategic coherence described in Chapter Six.

By acknowledging the ecological and dynamic dimensions in which the conditions of agency continually evolve, and by reasoning in accordance with the principles of coherence and the method of wide reflective equilibrium, prospective, purposive agents are thereby able to make informed strategic choices of moral and political principles. From this framework, Chapter Eight outlines the basic principles contractually designed to protect and enhance the global scope for evolving and maintaining the power of adaptive moral agency. Such principles are ultimately selected in terms of their rational and reasonable contribution to protecting and enhancing the mutual autonomy and well-being minimally required by PPAs. In effect, by continually safeguarding and fostering the development of these essential agency goods, this *organic* social contract thereby promotes the *adaptive* capacity for reflective autonomy which *moral* agency requires. This promotion of adaptive moral agency thus

becomes the rationale for a political conception of justice based on equal rights to the social goods needed by all PPAs. This is presented in terms of Rawls' and Scanlon's notions of rationality and reasonability, first raised in Chapter Three to defend the merits of the contractarian approach. In this context, issues concerning the legitimate political boundaries of the contract are also considered, which leads to the prescription for a more inclusive version to accommodate the evolved and evolving mutual concerns of adaptive moral agents.

Although the extensive rights and responsibilities outlined in Chapter Eight can be hypothetically legitimised through the model of an organic social contract, its genuine pragmatic value will consist in its own sociopolitical feasibility i.e. the model *itself* should also be sufficiently *adaptable* to enhance the agency needs of citizens here and now, in any social environment. However, organic selection pressures induce local social and material inequalities which must be continually rectified to preserve and enhance the conditions for co-operative moral action. Accordingly, in terms of its focus upon the dynamic interactions among agents' variable capability functionings in different localities, Sen's theoretical approach is identified as especially well-adapted to the political and economic tasks of promoting those co-operative conditions. To that end, the final chapter sketches an optimal economic model for measuring the fluctuations in the levels of agency goods in any social ecology, so that deleterious inequalities in those goods can be prevented or remedied as needed. Sen's index of capability functionings is presented as an ideal template for this task.

Yet, even with the robust support of Sen's model, biases in inherited social systems may continue to reward behaviours which discount or exploit the goods needed by other agents. Therefore, a motivational model is also outlined to support and sustain the moral disposition required for the development of mutual aid and trust. This model is conceived in terms of the ontogenetic processes described in Chapter Four, and aided by McClennen's dynamic strategy of 'resolute choice' and Gauthier's account of 'constrained maximisation'. By this account, the parameters governing the adaptive social choices of moral agents will most likely remain consistent with a broad version of reciprocal altruism. However, instead of being induced by organic selection pressures, which ultimately threaten to diminish the capacity for reflective autonomy, the acknowledgement of these evolutionary processes allows agents to autonomously direct the course of their own social and moral evolution. With this in mind, the contractual model of adaptive social choice sketched here is intended to serve as the basis for a subsequent work concerned with applying the model to finding concrete, long-term solutions to specific problems of distributive justice in societies at different stages and levels of social and economic development.

PART ONE:

THE EVOLUTION OF AGENCY

CHAPTER ONE

THE CO-EVOLUTION OF REASON AND AUTONOMY

It is now widely accepted that human beings owe their distinguishing physical characteristics to the force of natural selection. But the idea that much of human behaviour might also be accountable in these terms is often greeted with doubt and suspicion. To many, this idea appears as a threat to the autonomy that is believed to separate us from all other species. It is assumed that if our actions are the result of our genetic inheritance there is little or no room for autonomy. Freedom seems reduced to choosing between a variety of instincts or 'passions', neither of which are our own creation. But if the reasoning behind Darwin's theory is properly pursued, it would be a mistake to dismiss our perceived autonomy as an illusion. On the contrary, we can view our apparent capacity for self-control as evidence for its high selective advantage[11]. As such, an organism's autonomy can be conceived as the phenotypic expression of its power to adaptively modify its responses to environmental contingencies. The appearance of autonomy can then be more plausibly explained than by any other account and can be correctly classified as an essential characteristic of our own species, enhancing and maintaining the survival of its individual members. Thus conceived, the phenotypic evolution and maintenance of autonomous agency is a causal precondition for the ontogenetic growth of moral learning in each individual. In turn, the principles and practices which are normatively learnt and applied themselves become a cultural source of further social adaptations realised in the organism's phenotype. Accordingly, any contractualist account of the most stable reciprocal norms which may *theoretically* evolve among individuals in any society cannot afford to ignore the *practical* evolutionary preconditions needed for autonomous agents to become effective moral agents capable of selecting *adaptive* moral principles. With this ultimate goal in mind, it will first be appropriate to clarify the phylogenetic process by which autonomy naturally evolves in the phenotype of species generally, thereby illustrating the emerging synergy in the growth of reason and autonomy which has become a distinctive catalyst in the evolution of human agency.

The differing phenotypes that distinguish one species from another cannot be wholly determined by morphological changes. Often it is a divergence in patterns of behaviour that makes the crucial difference between life and death. Therefore, natural selection would be expected to favour the reproduction of a variation that displayed a tendency toward autonomous action. How such a variation might occur should be no more mysterious than any other. Any individual organism inherits a set of instructions that optimizes its functional capacities toward self-replication. However, self-replication alone clearly limits the potential for beneficial mutations to arise in the gene pool. Species that reproduce by making identical copies of themselves are dependent upon mutations alone to achieve and

maintain a viable fitness with their environments. But an organism that can combine its genes with those of another will produce offspring with a unique combination of genes every time. As the population increases, so does the degree of variation that might be useful in any given environment.

Sexual reproduction thus greatly enhances a species' survival prospects by generating many more variations than is possible by strict self-replication alone. As the latter process relies on a much lower frequency of random mutations, it is only effective for the simplest species that are already genetically well-adapted to relatively stable environments. The development of sexual reproduction is therefore crucial to a species' chances of evolving any sort of complex behaviour, instinctive or otherwise.

To avoid the presumption that instinctive behaviour necessarily precludes the exercise of autonomy, it will not do to simply accept that the two descriptions are mutually exclusive. It may seem contradictory to claim that a species' behaviour could qualify as instinctive and autonomous at the same time. Yet even basic breeding instincts exhibit elements of autonomy. Having noted sexual reproduction to be the precursor of all other complex behaviour, it can be regarded as the most primordial instinct. As such, it predates the development of any degree of rational thought that could effectively control it. The act of mating can be explained as a genetically programmed response to an environmental stimulus. Most animals have been observed to follow predictable patterns of mating behaviour that show little sign of control over their responsive range. But are interactions between genes and environment sufficient to explain why one pattern is followed rather than another?

Natural selection explains perfectly how a particular breeding strategy has become successful in a particular environment. Any existing mating behaviour will be ultimately traceable to a single, breeding pair. Either or both of these individual animals must have enacted a significant variation to the breeding programme, otherwise the innovation could never have been incorporated into the surviving phenotype. Any departure from the earlier pattern could have become significant only if it had an immediate effect on the survival of progeny.

In depositing its own eggs in the nest of a bird belonging to a different species, the cuckoo is following a genetically inherited strategy. But before such an activity could become a heritable feature of the species' behaviour, it must have been performed *against* its inherited 'instinct', if inheritance is limited to the transmission of genetic information. In other words, the bird that started this successful mischief was not at first genetically determined to do so. The genes that would favour the repetition of this strategy in subsequent generations cannot explain the contingency of its first appearance. Of course, the cuckoo's egg-laying behaviour, like that of the robin itself in feeding the alien hatchling, may simply be explained as a mistake.[12] Likewise, a genetic mutation from the species' life history can account for deviations from stereotypical behaviour, but that only begs the question of how that earlier mutation also produced that earlier response. Alternatively, it may well be the case that each female cuckoo 'learns' to identify the characteristics of its host species and nest, to which it returns to deposit its eggs as an adult.[13] But in that case, sheer genetic determinism of a patterned response is therefore untenable, as it does not permit the plasticity that is essential for a pattern to be learnt in the first place.

On the Galapagos Islands there lives a species of iguana that goes to extraordinary lengths to have its eggs buried under a layer of ash within an active volcano.[14] Here, it is the limitations imposed

by the environment that have clearly forced the iguana into a strenuous and high risk strategy. Nevertheless, while the island's terrain forcibly evoked the behaviour, it would not be enough to cause it directly. For, if among the first generations that had migrated to the islands, none had ventured up the slopes of the volcano, extinction would have been swift. But could the island be said to have *caused* their extinction any more than the volcano's presence had *caused* the survivors to bury their eggs in it? Certainly not. Like genes, environmental constraints on breeding strategies will explain why one strategy has been bred into the instincts of a particular species. But again, those constraints cannot deterministically ensure that the iguana will deposit its eggs in the volcano.[15] Similarly, it cannot be any ecological shortages or hazards that determined the cuckoo's striking departure from the instinct inherited from its ancestors. The fossil record proves that the environment does not design creatures with the best survival instincts. Rather, it selects those whose behaviour is more adaptive within the evolving ecosystem which is structured by both physical features of the landscape and the biological effects of other species' behaviour.[16]

These examples serve to show that instinctive behaviour, even when it is directed by the most primordial biological necessity, cannot be wholly determined by any combination of genetic or environmental pressures. For an organism to have *survived* those pressures testifies much more to its evolving flexibility of behaviour than to any fixed, rigid strategy. A species that cannot deviate from a set pattern has no chance of adapting to any environmental changes that threaten its livelihood. Conversely, a creature that has even the slightest degree of autonomous control will have the advantage of being able to explore alternative survival strategies. While many of these alternatives will turn out to be dead ends, some will dramatically enhance the prospects of a species' survival within a single generation.

(i)   Phenotypic Plasticity:
         The Baldwin Effect: Organic Selection

Given that some flexibility is essential for a species to deal with disruptions to ecological conditions, it follows that the greater the variety of responses within the phenotype that an individual inherits, the greater are its chances of surviving and reproducing. Armed with an array of responses to different conditions, such an individual will proliferate at the expense of those with fewer defensive or opportunistic skills and will take over the population. Gradually, instead of the environment exclusively shaping the organism's behaviour, the organism learns to manipulate the environment for its own benefit.

Throughout their lives, the vital activities of most organisms necessarily involve the capacity to identify and *select* appropriate foods, habitats, mating partners, strategies to evade predators etc. As many of these activities are "temporally mutually exclusive", the organism's genotype could not evolve by programming a set of rigid responses, but must instead favour the production of a range of phenotypes that allows for considerable plasticity in the adaptive "reaction norms" (Simpson) of each individual.[17] However, in seeking a solution to the question of how the "genetic assimilation" of such

adaptive responses might eventually occur, Waddington cited the work of the psychologist James Baldwin who had proposed the hypothesis of organic selection fifty years earlier.[18] In providing an explicit account of the mechanism whereby a specific phenotypic adaptation can subsequently come under the control of the genotype, while avoiding any appeal to Lamarckian inheritance, organic selection essentially involves the following sequence of events, as summarised by Hall:[19]

> (1) animals choose a new habitat or life style and adapt to it
>
> (2) mutations occur that favour the changes; and
>
> (3) selection favours those individuals best adapted to that new habitat or life style, and the adaptation spreads through population

While the Baldwin Effect is thus able to reconcile the organism's phenotypic plasticity with standard genetic inheritance processes governed by natural selection, developmental psychologists such as Oyama argue that attempts to reduce organismic adaptation to genetic factors alone is inadequate in failing to account for the "extraorganismic regularities" that must be maintained in the environment in order to support the emerging phenotypic traits.[20] In the very process of adapting to such environmental regularities, the organism effects changes in its surroundings. As such, the inheritance of phenotypic information need not become fully "assimilated" to genetic controls, particularly given that such complete fixation would effectively reduce the organism's capacity to respond to future environmental contingencies, leading to a *loss* of plasticity as Godfrey-Smith notes.[21] In that case, maintaining the *adaptive* capacity to learn must resist such complete "internalisation" in favour of maintaining variety in the *externalisation* of information, as our "symbolic species" has done through the use of language, culture and "niche construction".[22]

This process of organic selection is thereby thoroughly consistent with the substantial psychological evidence indicating that individual agents actively *select* social and moral principles and values, some of which are 'internalised' as habits or rules which then become external sources of information in the social environment. This evidence will be elaborated in the Chapter Four. Beforehand, in order to more fully appreciate the extent to which inherited cognitive and ecological factors are liable to either limit or enhance the degree to which autonomy and learning may evolve via both organic and natural selection, it will be instructive to consider a few examples in the phylogeny of different species.

(ii)   Autonomy and Learning:
         Species-Specific Examples

Nowadays much animal behaviour that had long been attributed to 'blind' instinct has since been shown to be learnt. In the past, a species' behaviour was routinely labelled "instinctive" often on the groundless assumption that its members lacked the level of intelligence required for effective learning. But as the behaviourists' experiments later showed, even a 'bird-brain' can direct the performance of tasks that would be impossible without an innate capacity to learn.[23] As Skinner was able to elicit similar, predictable responses from a range of species including humans, he concluded

that learned behaviour generally did not require any imputation of an active self-aware mind.[24] The capacity to form mental representations of objects did not even seem to be a contributing factor in the learning process. All that was required for successful learning was the purely sensory awareness sufficient to recognize the appropriate environmental cue to perform a particular act.

The behaviourist model conforms well to the Darwinian view, in so far as it demonstrates that an organism's actions can be reinforced by repeated exposure to environmental effects.[25] By showing how an 'instinct' can be aroused, modified or even extinguished by these variable influences, the adaptive capacity of an organism is seen to be synonymous with its learning capacity. But if behavioural changes were only ever passive, predictable responses then autonomy would indeed be superfluous. However, even the most primitive and despised insects could not function without a modicum of autonomy to guide them. To understand how autonomy has evolved to function as a central control mechanism, consider the most basic ability common to most organisms - locomotion. The constant search for food and mating opportunities requires varying degrees of mobility. The fly that repeatedly attacks your barbecued sausage succeeds by initiating a series of intricately executed manoeuvres. But none of these infinitely variable responses can be programmed or learnt in advance. To negotiate a successful landing on the sausage as it enters your mouth, the fly must make an informed decision about the likely speed, direction and distance of the target in order to adjust its immediate flight path. Equally, it must continually survey its surroundings for any obstacles that appear in its path. As Warren notes, insects' orienting movements cannot be adequately explained as induced responses to a given set of stimuli.[26] In this case, the scent of the sausage is undoubtedly the stimulus for the fly's approach, but the scent alone cannot determine the specific course the fly will negotiate, nor precisely when it may respond. The tree that blocks the most direct route may be avoided by veering to the left or right or by going between this or that branch. Suddenly from beneath the tree wafts the smell of a fresh dog faeces. Decisions, decisions!

Insects' motor skills could not have evolved as efficient guidance systems unless they were equipped to respond to a range of sensory data. If evolution had only allowed them to learn by matching a rigid, stereotyped response to a specific environmental stimulus, they would have been severely handicapped. Adaptation to environmental changes could only be effected by sheer luck, which is evidently not the case in insect behaviour. However, it is not any newly developed behaviours that make insects so hardy, nor any intelligence.[27] Their success is due to their obedience to laws of their own design i.e. laws of autonomy.

An ant guides its own actions by following the communal signals it shares with its fellow ants. At first glance, this mindless obedience to the societal norm appears as the antithesis of autonomy, since no ant displays any inclination to defect. To our understanding an ant colony resembles a totalitarian dictatorship. Yet, as Wilson has pointed out, the colony contains no incentive or coercive elements that might ensure conformity. The activity of a colony can only be construed as "the summation of a vast number of personal decisions by individual ants."[28] The fact that those decisions invariably coincide need not signify a lack of choice, as is usually assumed, but may be the living expression of a categorical imperative produced by natural selection.

The ant's aptitude for organising and regulating its activities for the benefit of the colony is fully inherited. Instructions for the performance of these tasks are encoded in the genes and are translated into action by the formation of rules. Remarkably few of these genetic algorithms are needed to enable an ant to function efficiently. Recent computer models have shown that the application of four simple rules is enough to direct the foraging behaviour of an entire ant colony.[29] In so far as those rules are optimally designed for foraging under most conditions the species is likely to encounter, then any alternative strategies that may have been explored will be rapidly extinguished. In game-theoretic terms, an evolutionary stable strategy (ESS) has been reached where there is no value in pursuing alternatives. When all the other ants are following the same rules with maximum efficiency and prosperity for the colony, any ant that deviates from those rules risks losing the protection provided by the colony.

Like flies, ants are also attracted by following scents. Before long your picnic leftovers are swarming with them. But none of the ant's genetic algorithms include detailed up-to-date maps of the picnic grounds. The precise location of a cold chicken leg is an unpredictable variable in the landscape. Consequently, no specific algorithm has evolved that is dedicated to harvesting this resource. Instead, the colony's survival depends upon keeping its resource options open. So the rules that guide foraging have not been entirely fixed by natural selection. A species that developed a taste for only cucumber sandwiches might survive and even prosper for a time in some parts of Britain. But dependence upon such a specialised diet is a high risk strategy not likely to be rewarded in the long run. Therefore, it pays to evolve algorithms that allow for flexibility in tracking food sources. After all, it is the energy converted by food that sustains life, not any particular food.

In the competition for resources it pays to be an omnivore. But to become an omnivore requires experimentation and exploration. Biologists typically reduce the ant's foraging behaviour to its strategy for laying trails with chemical messages that alert other ants to the presence of food. The act of following pheromonal trails thus constitutes the primary rule of foraging behaviour. However, having the equipment to follow a trail does not provide the recipe for laying one. Every ant must also be ready to lead its fellows to its discoveries. To be a trailblazer requires an inclination to explore the environment without a map. As a new generation leaves its nest for the first time there are no trails to guide them.[30] A trail is first formed as the result of each ant's singular excursions. If every ant blindly followed the first one out of the nest, the fortunes of the colony would be placed at risk.

The exploratory excursions of ants must begin as random wanderings guided only by sensory information. Once an ant encounters a possible food source, it can inform the other members of the colony and follow one of their trails back to the nest. The scope of this foraging ant's autonomy is, of course, restricted to completing a task for the benefit of the colony. But its foraging could not be successful without the genuine autonomy that is required to pursue alternative paths.

Closer to our own species' evolutionary heritage is another much maligned survivor - the mouse. The information processing of a mouse's brain evolved in the same direction as our own expanded mammalian brain. Algorithms of plasticity have evolved in the rodent brain that not only permit exploratory behaviour but develop it directly through learning. This has been amply demonstrated in the classic maze experiments that test learning capacities. In a notable experiment the

white-footed mouse learns to run through a maze to reach a reward.[31] What surprised the researchers was that even after it had learnt the correct route, the mouse continued to make occasional excursions along different routes. The evolutionary advantage of this behaviour is as clear for the mouse as it is for us: it pays to pursue alternatives rather than relying on one course of action. In the event of rewards diminishing from one strategy, others have also been learnt and remain open alternatives. Thus, the organism's capacity for self-control also entails the capacity to recognise and respond to alternative options. Our species has learnt to apply this rule of autonomy to the greatest advantage. So much so that we could not function without it.

The autonomy observed in the activities of an ant, a fly or a mouse may appear qualitatively different from that displayed by humans. In particular, the kind of autonomy that would permit the evolution of morality is evidently lacking in all other species. Where is the ants' police, the flies' High Court, or the Maze prison for mice? The fly that spoils your food cannot be prosecuted for wilful damage as whatever 'will' it may possess is not subject to autonomous constraint. Kant's conception of the "moral law" draws this crucial distinction between a will subject to determination by instinctive desires, as opposed to a will guided by reason alone. For Kant, the autonomy required for moral actions can only be properly attained by the intervention of reason.

Kant proceeded upon the assumption that reason is the preserve of humans. As long as there was little or no scientific knowledge of animal behaviour this anthropocentric view may have been plausible. Now there is a wealth of evidence to show that those species closely related to humans employ the same kind of concepts and categories that Kant ascribed to human thought.[32] Many primate species are particularly adept at learning in ways that cannot be reduced to conditioned responses. These animals have a talent for invention that can only be realised through the understanding of certain rational concepts and laws. Until recently, tool-making was often cited as evidence of an ingenuity peculiar to humans. But it is now known that chimpanzees, for instance, use twigs as tools for extracting termites from hollow logs, stones for cracking nuts.[33]

The discovery of these crude tools signifies a great leap forward in the power to process information. For it marks the development of inductive, instrumental reasoning which is the hallmark of our own species' prodigious inventiveness. To form the conception of an object as an instrument for producing a desired effect implies a basic 'intuition' of the concepts of space and time.[34] An accurate perception of space and time is vital for most organisms, as the fly's behaviour showed. But this perception need not be accompanied by any conscious, conceptual representation in the fly's brain. The autonomy that controls its flight pattern is contained entirely within its genotype. Natural selection has gradually fitted the expression of its autonomy into directing a limited repertoire of activities such as flight. As the fly has no need to learn the rules of aerodynamics, it can function perfectly without any conceptual 'knowledge' of those rules.

By contrast, primates have developed minds capable of consciously representing objects perceived in space and time. An imaginative consciousness goes beyond the simple observation of objects in space and time. By reflecting upon the recurrence and proximity of those objects, the imagination can conceive causal 'necessary' relations between them. On this basis inferences can be drawn, probabilities assessed, actions planned. By noting a necessary relation between flight and the

motion of wings through a continuum of space and time, humans have constructed flying machines. Yet the same instrumental reasoning that can send a vehicle into unexplored regions of outer space also informs the chimpanzee how to catch termites with a twig. The fact that our ancestors discovered the possibility of capturing fire with the same tools does not mean they reasoned any differently from other modern primates. Experiments have shown that some chimps can at least learn to tend a man-made fire.[35] They may well lack the 'intelligence' to envisage the culinary possibilities of roasted termites-on-a-stick or baked bananas. But their survival without the use of fire only shows that it never became necessary to develop this intelligence. As our early ancestors migrated from the forests to open savannah, they had to change to a meat-eating diet just to survive. If they had not made use of fire for warmth and cooking, they might not have adapted to this new environment with much success, in which case we might not be having barbecues today.

"Necessity is the mother of invention" is an old adage that perfectly captures the essence of instrumental reasoning. It also explains the selective advantage of being innately equipped to manipulate objects and events in relations of means and ends. For without an intuition of necessity our minds would never be able to conceive of any possible function for a given object. All phenomena would remain devoid of meaning and value for us. But the instrumental significance by which we now classify virtually everything would have proceeded gradually, in response to the process of natural selection. Our initial discoveries would have been more like that of the chimpanzees. The use of fire and crude implements can be stumbled upon by trial and error alone without any abstract concept for the principle of necessity.

The chimp's experimental discovery of the twig as a means of catching termites is driven by the necessity of finding food to survive. The search for food need not be provoked by immediate hunger. Like humans, most animals have learnt the habit of hunting frequently as a safeguard against starvation. Not because they have any knowledge of the danger of death. Natural selection enforces the biological imperative of survival for reproduction via the sensory mechanisms of pain and pleasure. So it is that needs and desires evolve to motivate the organism toward this ultimate goal. But in order to act upon those needs and desires, the organism must be able to initiate and control its actions in the right direction. This is where autonomy has its function. Simply feeling a need is not sufficient to cause any response to it. The process of satisfying a need does not come about like a reflex action.[36] In a single mind, any number of competing needs may be demanding action. Similarly, there may be a variety of means available for meeting a given need. Unless there is only one prospective food source nearby, the animal will have to make decisions. If there is a nest of termites here, a bunch of bananas there, some juicy leaves over here etc. the chimp may well choose the tastier one, e.g. the bunch of bananas. But, then again, he has to use more energy to reach them. He looks at the leaves again. Mmm, they do look juicy. But there are two bushes to choose between, both bearing equally juicy leaves.

As these examples show, the capacity for autonomous judgement and learning is an evolved function of the organism's biological needs within a particular environment. However, in the case of highly intelligent social species, such as chimpanzees and humans, those needs must also be negotiated in a *social* environment where co-operation among individuals is crucial to survival. Therefore, the next chapter will show how moral and political systems evolved to ensure that basic goods may be distributed to the common benefit of individual group members.

---

[11] Dennett, op. cit., p.367

[12] Richard Dawkins, *The Extended Phenotype: The Gene as the Unit of Selection* (Oxford, W.H. Freeman, 1982), p.54

[13] Ibid., p.94

[14] R. Perry, *Galápagos* (Oxford; New York, Pergamon Press, 1984), pp.166-167.

> See p.171: 95% of females travel distance up to 15 km – ascent of 1000-1300m to rim of volcano, descent of about 900m. Journey to rim takes 4-14 days (9 av.) Descending to egg-laying site, digging nest, laying eggs and returning to rim takes another 20-33 days (23 av.) Round trip for nesting – 32 days av.

[15] Cf. Sterelny's account of Goannas on Krakatau in "Darwin's Tangled Bank" in *The Evolution of Agency and Other Essays* (Cambridge, Cambridge University Press, 2001), p.161

[16] Ibid., pp.162-164

[17] Gordon H. Orians, "Behavior, Ecology, and Evolution," in *A New Century of Biology*, ed. W. J. Kress & G. W. Barrett (Washington, DC, Smithsonian Institution Press, 2001), pp.96-97

[18] David J. Depew, "Baldwin and His Many Effects," in *Evolution and Learning*, Weber & Depew eds., p.16; pp.3-4: Note Lloyd Morgan's & Osborn's two independent proposals all in 1896

[19] Brian K. Hall, "Organic Selection and Genetic Assimilation," in Weber & Depew eds., *Evolution and Learning*, p.143

[20] Susan Oyama, "On Having a Hammer," in Weber & Depew eds., *Evolution and Learning*, p.178

[21] Peter Godfrey-Smith, "Between Baldwin Skepticism and Baldwin Boosterism", in Weber & Depew eds., *Evolution and Learning*, pp.55 & 57-58

[22] See Terrence W. Deacon, "Multilevel Selection in a Complex Adaptive System: The Problem of Language Origins," in Weber & Depew, *Evolution and Learning*, Ch.5 and his *The Symbolic Species: The Co-Evolution of Language and the Brain* (New York, Norton, 1997). See also F. J. Odling-Smee, K. N. Laland & M. W. Feldman, *Niche Construction : The Neglected Process in Evolution* (Princeton, N.J.; Oxford, Princeton University Press, 2003)

[23] Leslie Rogers, *Minds of Their Own : Thinking and Awareness in Animals* (St Leonards, N.S.W., Allen & Unwin, 1997), pp. 36, 58, 72, 85-88, 128, 160-161

[24] See *Behaviour of Organisms* Ch.1 for behaviourist methodology; *Science and Human Behaviour* pp.63 -71 for pigeon experiments; *Beyond Freedom and Dignity* for wide-ranging application to human behaviour and society

[25] Dennett, op. cit., p.374

[26] E. C. Goldfield, *Emergent Forms : Origins and Early Development of Human Action and Perception* (New York, Oxford University Press, 1995), pp.193-4

[27] Edward O. Wilson, *In Search of Nature*. (Washington, D.C., Island Press, 1996), p.69

[28] Ibid., p.68

[29] See P. Coveney, & J. R. L. Highfield, *Frontiers of Complexity: The Search for Order in a Chaotic World* (London, Faber, 1996), p.249 for four "Rules of ant life": If you find food - take it home marking a trail; If you cross a trail and you have no food – follow the trail to the food; If you return home with food – put it down and return up the trail; Otherwise – wander at random looking for food

[30] Ibid., p.248

[31] Waller, *The Natural Selection of Autonomy*, p.7

[32] Immanuel Kant, *Immanuel Kant's Critique of Pure Reason*, trans. Norman Kemp Smith (London Macmillan, 1961) B95 & 106

[33] Frans De Waal, *Good Natured : The Origins of Right And Wrong in Humans and Other Animals* (Cambridge, Mass ; London, Harvard University Press. 1996), p.72

[34] Kant, *Critique of Pure Reason*, B34-36

[35] W. C. McGrew, *Chimpanzee Material Culture : Implications for Human Evolution*. (Cambridge, Cambridge University Press 1992), pp.11, 148, 229-230

[36] Cf. Dewey's critique of "The Reflex Arc Concept in Psychology," in *The Essential Dewey* vol. 2, ed. L. Hickman & T. M. Alexander (Bloomington, Indiana University Press, 1998), pp.8-9

CHAPTER TWO

THE EVOLUTION OF MORAL AND POLITICAL SYSTEMS

Fortunately, unlike the proverbial Buridan's ass, animals do not normally starve from indecision. As shown in the previous chapter, natural selection ensures the survival of animals that are fitted with a sufficient degree of autonomy to satisfy their biological needs. A range of emotions serves to alert the animal to those needs. The same process occurs in the human animal. A child who is bitten by a dog quickly learns the necessity of avoiding it in future. It is reminded of this necessity by the feeling of fear. Fear is therefore a very effective means of preventing an act. Conversely, desire functions as a means of provoking an act that will give pleasure. To receive the pleasure, it is necessary to perform the act. Here we have the foundations of moral sensibility as a means of controlling human behaviour. Accordingly, this chapter will be largely concerned with providing empirical evidence to indicate that humans have simply adapted these inherited biological control mechanisms to the demands of living in extended groups. With the aid of instrumental reasoning, these demands can be met more effectively by rewarding activities that produce resources and punishing those that deplete them.[37]

For the same reason that the human body evolved into an apparatus well suited to the environmental demands placed upon it, the human mind cannot have developed in an arbitrary fashion as if it were immune from the need to adapt. As Kant noted, reason is by nature a practical method of interpreting and acting upon the world.[38] The evolutionary value of the abilities to categorise objects in space and time, and to infer causal relations between events would have conferred our ancestors with enormous advantages over other species competing for resources. Needs can be met much more efficiently by manipulating those resources in new ways that yield benefits at a much lower cost to both the individual and the group. The use of fire and agriculture are two classic examples in our history of the successful application of inductive reasoning to basic problems of survival. As many species survive and reproduce most effectively by living in groups, the evolution of human societies and cultures owe their success to co-operative efforts at meeting the needs of individual members. Each person's livelihood thereby became bound to the fortunes of the group. Thus, as a system of rules and norms directing human social behaviour, the concept of morality arises as the best means of ensuring that each member contributes to the group's survival strategies.

It should be noted that this evolutionary basis of morality need no longer be regarded as a speculative 'just so' story. In fact, there is now a wealth of empirical evidence to support a common evolutionary functionalism in the generation of moral concepts and norms throughout all surviving cultures, at least insofar as 'morality' is primarily concerned with regulating the co-operative behaviour of individuals. To be sure, the prehistoric origins of moral behaviour make it impossible to establish conclusive theories explaining *all* its conceivable normative content in *all* prehistoric communities. But the impossibility of direct observation and experience of prehistoric cultural life need not prevent substantial discoveries from a range of different sources. Thus, a scientific interpretation of the prehistoric mind cannot justifiably be limited to formulating hypotheses about social behaviour solely by testing the possible functions of ancient artefacts and the remnants of ritual acts. If the primary function of morality does, in fact, consist in promoting the mutual benefits of social co-operation, then support for such a hypothesis can also be obtained from the investigation of moral codes and practices in those extant bands or tribes whose subsistence activities are demonstrably similar to those common in prehistory. The social behaviour of closely related primate species can also provide an additional source of information by which to test the phenotypic stability and strength of the cognitive mechanisms which continually enforce co-operation. For example, while reciprocity is commonly practised throughout the troupe, the dominant chimpanzees also use aggression to suppress conflicts among troupe members.[39] Finally and most importantly, evidence from all of these historical and contemporary ethological sources can be tested against the results of recent research in cognitive and social psychology and studies in economic rationality. This research will be presented in Chapter Four in connection with the ontogeny of moral learning.

Widespread archeological evidence of co-operative hunting among hunter-gatherers in the Middle to Upper Paleolithic can be inferred from the combined presence of stone-tools, fossilised bones used for butchering large animals and weapons such as spears, darts and harpoons.[40] The Upper Paleolithic era is marked by a transition from subsistence foraging in small bands to larger populations sustained by intensive and co-ordinated provisioning strategies, leading to the production and trade in

the first "luxury goods" and symbolic artefacts.[41] These adornments, such as pendants and beads made from ivory and sea shells acquired from foreign locations, indicates the extension of co-operative exchanges and intermingling of tribes.[42] The production, display and trade of similar ornaments among members of existing foraging societies can provide a plausible model for interpreting their original social function. Accordingly, it is reasonable to assume that both the egalitarianism practised by modern nomadic foragers and the social stratification observed in existing sedentary farming communities provide an insight into functioning of the prototypical moral systems likely to have governed social behaviour in the Paleolithic.

In keeping with the cost/benefit mode of planning material goals in the world moral reasoning is similarly directed and constrained by the need to bring about or maintain a particular state of affairs. In the case of our primitive ancestors these needs and conditions were the most basic 'goods' of food and shelter. The extent to which each person shared in these goods would have been a convenient measure of their cost to the group, while their role in providing such goods would give them a creditable value.[43] From that point on, the concept of justice emerges in the assignment of values to social roles on the basis of productivity, bringing those in debt to account, while rewarding successful leaders with privileges and higher status.[44]

Due to the overwhelming reliance on fossilised material, archeological evidence for the emergence of such hierarchical social distinctions in the Paleolithic is limited to the numerous beads, pendants, bracelets and red ochre found in burial sites.[45] The most recent archeological findings provide increasing evidence for trading in such material goods as early as 100, 000 years ago, indicating the crucial role of reciprocity between groups as a civilising influence in social evolution.[46] However, there is substantial and clear evidence from more recent prehistoric eras to show the economic origins and function of the concept of justice. For example, the ancestors of the existing Wanka population in the central Andes have left a variety of artefacts which indicate the development of a social hierarchy following the intensification of agriculture and the specialisation of productive tasks.[47] Excavations of the earliest deposits from 200 B.C. have revealed mostly utilitarian goods such as mortars and pestles, serving bowls, storage and cooking jars made in a ceramic style common among distinct settlements throughout the region, indicating regular interaction between groups. The dispersed presence of obsidian also shows the occurrence of long-distance trade.[48] The earliest symbolic artifacts are ceramic figurines depicting warriors holding knives and trophy heads and figurines of camelids (a species of llama) "many of which are smiling and very fat (pregnant)?"[49] Later deposits show evidence of a gradual transition to increased production and processing of plant and animal resources, with a corresponding decrease in both the production of figurines and little trade with outsiders. The final phase in the Sausa's cultural development is marked by increasing production costs and benefits differentially distributed among households, as indicated by the differences in the size and construction of stone houses and enclosed patios and in the household's quantity of status items such as cloth and metal.[50]

As Hastorf notes, insight into the significance of these artefacts is also available from the consistency of ethnohistoric accounts related by the descendants of those prehistoric populations.[51] Likewise, knowledge of the significance of wealth displays, social roles and exchange rites both within

and between the descendants of various foraging tribes can also be used to trace the continuity of function in ancient and modern political and moral systems. As Megarry notes, despite some ecological disparities in using modern foraging societies as models of prehistoric social organisation, subsistence societies, by definition, necessarily experience "current needs" which occur in all geographic locations.[52] Accordingly, an investigation of the moral concepts and practices of existing foraging and agricultural societies can also serve as evidence for the evolutionary stability of their socio-economic functions. This is amply demonstrated in the ethnographic data drawn from two tribal societies occupying radically distinct environments in different eras: the !Kung of the Kalahari in the 20th century and the Iroquois of 17th century North America. Prior to the territorial incursions of white settlers in the 1970s, the !Kung were famed as the staunchest of self-sufficient egalitarians, who "abhorred and feared conflict and who placed a high value on peace and harmony".[53] In such a society, there is evidently little or no prestige attached to positions of leadership.[54] This is typical of small foraging societies where there is no incentive for individuals seeking to acquire more than an equitable share of resources, when such goods are regularly and freely accessible to all members of the group as needed. By contrast, among agriculturalists like the Iroquois, the continual surplus production of goods does indeed provide a powerful impetus for the acquisition of status and authority. Among the Iroquoi, men achieve prestige through success as warriors and in their personal prowess and industriousness in hunting, fishing and providing feasts.[55]

Generosity in communal provisioning is a normative characteristic of leadership in both egalitarian and rank societies. However, as its exercise in an egalitarian system does not result in the endowment of any special privileges which would induce competition and selection for such generous acts, there may seem to be no place for the rewards and punishments which are elementary features in the evolution of moral systems. Closer inspection of the *actual*, as opposed to the ideal conception of egalitarian practices reveals its dependence on co-ercive sanctions such as gossip, shaming and ostracism that are frequently employed to suppress any self-seeking behaviour and ambition.[56] In this sense, as Boehm argues, egalitarianism in tribal groups is achieved in practice, not by the absence or relaxation of social controls, but is intentionally maintained by the regular operation of "antihierarchical" sanctions which, in the case of the !Kung, place a negative value on individual domination, arrogance and "stinginess".[57] Thus, rather than being the expression of some innate disposition for "kindness, generosity, affection, hospitality", in such a mode of life those 'virtues' are more properly described as "necessities for survival…the *essential properties* of the social structure, not just as values and norms; because lacking such (structural and functional) social attributes, the fluid, loose social system would collapse."[58]

It is precisely in these terms that such virtues, norms and values are the evolutionary product of a perpetual process of artificial selection and co-operative cultivation among the great majority of tribal members. As such, this process also accounts for the manner in which social castes evolve, further reinforcing the structure of moral value to the point where there is such a general consensus of understanding "good" that it appears as part of the natural world itself.[59] This is what Mackie refers to as the "error theory" of attributing moral beliefs and values with an objective status, independent of human interest or design, as if they belonged to the "fabric of the world", as Hare suggests.[60] In the

conduct of our daily lives, we all tend to subscribe to this theory out of sheer psychological necessity. However, before examining the psychological evidence for a primarily prescriptive functionalism in moral judgement, it will be appropriate at this point to briefly raise a logical question which is often seen to challenge the legitimacy of naturalistic approaches to assessing the validity of normative ethical judgements.

Ever since Hume, the distinction between facts and values has plagued all attempts to draw prescriptive lessons from claims about human nature. The mere fact that our species has evolved to depend upon reason is taken as irrelevant to the question of whether such a faculty is morally desirable. Moore's "naturalistic fallacy" makes the same point arguing that, for any human trait, it always remains an "open question" whether it is good to have that trait.[61] On its own, the question of what constitutes a good act or the 'good life' is meaningless and open to any number of equally speculative answers. However, it remains an open question only so long as it is devoid of any reference or goal. But an intentional act, by definition, always refers to a goal. Once the goal has been posited the question of value arises immediately.

Hunger reminds us that we need sufficient food if only to supply the energy required to find more food. Each time you eat you implicitly answer the 'open' question: "To be or not to be?" The fact of life is thereby given a necessary value, in that it is necessary to live if only to satisfy further needs or desires. The desire for food and sex has evolved to ensure that we value life enough to reproduce it. They are biological facts which have an objective, necessary value for many of us, for we would not exist without them. The question of whether or not we ought to eat or reproduce, as opposed to refraining from either activity, is therefore extraordinarily vacuous, when such goals are undeniably essential for our own continued existence as human agents. If we did not already sufficiently believe that such activities had a necessary value we would not engage in them as often as we do. Of course, it can never be demonstrated that life is necessarily preferable to death, or that anyone ought to live merely because they do live. Likewise, this objective need for physical sustenance should not be taken to imply that one is somehow obliged to eat on any *particular* occasion, as if the mere experience of hunger cannot be suppressed or overridden by other intentions. Clearly, innumerable occasions may be cited where eating may not be desirable, as in the case of maintaining a hunger strike, or may even be positively inimical to one's health as when only putrid or poisonous foods are available. Similarly, from the perspective of their own *personal* well-being, there is no biological imperative for any given individual to ever produce any offspring. However, as long as human beings can only come into existence via the reproductive efforts of other human beings, then those efforts must have an objective value for any individual who values their own existence. Therefore, in terms of their general necessity, it can hardly be considered a fallacy to believe that certain facts can be perceived as being good or bad. The assertion that "eating is good" is not analytically necessary as "1+1=2" is. Rather, its necessary value consists in the activity being an *essential precondition* for the success of the organism's continued functioning. Thus, for a being who wishes to continue functioning effectively, in this case as an autonomous human agent, it is an objectively valid belief to hold, given the fact that we are beings who must eat to function as such. In this sense, for such a person the general belief that "eating is good" obtains its *objective* validity from the empirical fact that the consumption of food is necessary

for their survival and well-being. Accordingly, it is a proposition that obtains its *truth* value from its correspondence with the *actual* biological conditions that must be present for that person to continue to exist.

It is crucial to note that this objective validity in the use of an evaluative term such as 'good' does not imply any attribution of goodness which is independent of the particular individual's interests or motives, as intuitionism would have it. 'Good' in this context should typically be understood as an abbreviation of 'good for' a specified function, which may be more or less requisite for that function. In this case, as eating becomes *absolutely* necessary for any person who wishes to survive, "eating is good" is indeed an objectively valid proposition if 'good' means 'essential for survival'. At the same time, given that this objective validity only arises as a result of the individual's own personal interest in their survival or well-being, that *subjective* appraisal must also be present for the *objective* evaluation to be coherent. In other words, the objectively valid belief that eating is good for survival can only be valid for a being capable of appreciating the instrumental value of the causal necessity that links 'eating' with 'survival'. As such, even a hunger striker's determination to *refrain* from eating is itself coherent precisely *because* it is premissed upon the implicit acknowledgement that eating *is* good (necessary) for survival. Likewise, a preference to remain childless would lack epistemic coherence if the person holding that preference failed to acknowledge the necessity of taking measures to avoid pregnancy.

Accordingly, a belief can indeed be subjectively valid by the extent to which it is epistemically consistent with other beliefs concurrently held by the same individual. By the same token, a belief can be objectively invalid in being inconsistent with known facts, as would be the case if someone were to deny that eating is essential for survival in human beings. As the objective validity of a belief is ultimately determined by its coherence with observed facts, such beliefs are falsifiable by those facts. By contrast, as a moral belief or value judgement can be subjectively valid if it is not inconsistent with a person's other beliefs or preferences, it makes no objective claims which are amenable to error. Nevertheless, as noted in the above examples, value judgements (moral or otherwise) *are* open to incoherence if they fail to acknowledge the *objective* means necessary for their realisation.

These normative standards of coherence will be discussed in much greater detail in Chapter Six, while the question of fact/value, is/ought relations and the 'naturalistic fallacy' will be directly addressed in Chapter Four. At this stage, these issues are raised in support of the meta-ethical claim that ascriptions of value, including moral and political value, can best be explained in terms of their functional necessity within an *evolving system* of beliefs and practices that are largely based on principles of inductive reasoning. For this reason, the adaptive function of any particular belief, desire or value can only be appreciated by an inductive analysis of its contribution within that system. To assess whether or not a moral belief is necessarily 'good' for the well-being of individuals in a given environment, the value of that belief must be judged by its actual effects in that environment.

Hume's scepticism over the epistemological justification of necessity led him to be equally sceptical of its role in justifying moral beliefs. As necessity is essentially an inferred relation between ideas or observations that are perceptually distinct, it cannot be used to justify any beliefs in facts.

There is no observable "necessary connection" between events themselves, so the concept of necessity has no real objective status, let alone any moral import.[62] However, it does not matter that the human brain may have evolved to perceive necessity as objective when it is more likely induced by human interest. It only means that we recognize that all value is determined by interest, rather than disinterest. In other words, we cannot regard our lives with a disinterested, valueless "view from nowhere".[63] Everything we do implies the preferential selection and evaluation of information. It is our common experience of the necessity to act that gives us "good cause" to reason from "is" to "ought".

Prescriptive terms like "ought", "should", "must" are frequently used to convey evaluative judgements about the "best" route to take, the "right" tool for the job etc.[64] Long before any distinctly moral category of thought had developed, language enabled humans to give practical advice to each other, without having to learn it directly through personal trial and error. For example, a single member of a nomadic tribe can give precise directions to enable others to locate a fresh carcass or waterhole, thereby dramatically reducing the time and energy invested in hunting. Directing a nomadic stranger to the nearest fast-food restaurant would involve the use of evaluative and imperative terms: "You must go left at the next corner ahead. Then you should turn right at the crossroads and stop at the golden arches. That would be your best bet." It is important to note that the imperative here is not intended to be categorical. As with any advice, it is conditional upon the express wish that motivates it. In effect, it is a hypothetical imperative because it is valid only if it is epistemically warranted by the corresponding goal that motivates it.

In the example, the directions given are 'right' for the particular purpose of finding the nearest fast-food restaurant. Assuming the stranger's express wish to satisfy his hunger immediately, the advice is good because it details the most efficient means of realizing that goal in the circumstances. All this is implicitly understood and assessed when making any practical judgment. Fundamentally, it is the utility of means in relation to ends that underlies the concept of good. So directing a hungry stranger to a restaurant is regarded by both parties as a good deed only because it is useful to the stranger's interest. In all such cases, goodness is synonymous with utility.

Altruism, broadly defined as the moral concern with satisfying another person's interests, does not need any special explanation. It does not involve any significant shift in value judgments. The utility of aiding another is a perfect example of an evolutionary adaptation.[65] All that is required is the intelligence to realise that the other members of the group have a similar interest in their own survival, as you do in yours. Conceiving a "theory of mind"[66] that explains and predicts the behaviour of others is a crucial step in the development of moral thinking. But it is one that can be readily taken by a mind that has already developed a high degree of self-awareness, for it only involves recognising that the behaviour of others is similarly self-motivated. Reviewing the psychological research on the cognitive elements involved in motivating altruistic behaviour, Batson notes that the "perception of another in need" appears to require the following three conditions functioning simultaneously at a certain threshold:[67]

  (a) a perceptible discrepancy (real or apparent) between the other's current and potential states on some dimension(s) of well-being

  (b) sufficient salience of these states, so that each can be noticed and a comparison made

(c) the perceiver's attention being focused on the person in need, not on the self or some
other aspect of the environment

The "dimensions of well-being" noted in (a) "include being free from unpleasant states -physical pain, negative affect, anxiety, and stress – as well as experiencing pleasant states – physical pleasure, positive affect, satisfaction, and security."[68] While the reliable operation of these combined cognitive conditions may be necessary for the perception of another's need, the research suggests that they are not sufficient to arouse genuine altruistic intentions. Psychological conditions which are both necessary *and* sufficient to motivate an altruistic response require that the perception of the other's need should be grasped by imaginatively adopting that person's own perspective. By empathising with the other's affective state, their perceived need can be imagined more keenly, making the desire to satisfy that need urgent enough to generate action on their behalf.[69]

Therefore, in accounting for the evolution of a mutual moral concern, once an individual has grasped the fact that the other members of the group are ordinarily motivated by these same psychological mechanisms which evoke that concern, then survival is conditioned by the satisfaction of mutual interests. So it is that norms regulating group living first evolved as a more efficient means of ensuring the survival of those individuals who share the same genetic self-interest. The habit of investing resources in the welfare of those 'nearest and dearest' is characteristic of all social animals. By investing in the success of its immediate family, a single organism effectively ensures the propagation of its genes. Thus the process of "kin selection" is the direct result of the natural selection that operates on individual organisms.[70] So there is no need to invoke a questionable process of group selection to account for the apparent selflessness that an animal shows for its offspring or siblings. For they all represent additional vehicles for the survival of its genes .[71]

In examining the phylogenetic evolution of altruism in humans, it is important to distinguish between different levels of selection that are necessary and sufficient to account for such behaviour. In Dawkins' theory, precisely because it is an organism's genes that are the basic unit of selection and not the living organism itself, it is not really correct to speak of 'selfishness'.[72] As the very idea of natural selection is to show how organisms exist without any design for their development, an organism's genes cannot have any 'interest' in their reproduction. For the same reason, it would be absurd to suggest that humans have children out of the selfish desire to immortalise their genes. The "selfish gene" is best understood as a metaphorical short-hand way of stressing that it is really only genes that reproduce themselves, not people.

At the level of individual behaviour, selection will also be operative in a corresponding psychological domain, which may or may not conflict with selection processes at the genic level. Therefore, while the behaviour of closely related individuals is naturally generous and supportive, those psychological dispositions may still be sufficiently supported at the genic level in evolutionary terms, given that they all act on behalf of the same genetically inherited interest. Modern families extend the same support and indulgence even to in-laws for just the same reason, thus allowing an individual's reproductive success to be measured in terms of the "inclusive fitness" of relatives.[73] In conformity with the principle of kin selection, most individuals move within a circle of familial relations who generally look after each other's interests without seeking any reward.[74] Nevertheless,

ethological evidence shows inclusive fitness to be a rather weak impetus in reducing conflict among kin.[75] In any case, as noted above, acts of generosity among kin may be genuinely altruistic in the psychological domain, yet remain driven by a common genetic interest produced by genic selection. Therefore, while it is indeed vital to distinguish between evolutionary altruism and psychological altruism, fitness measured in terms of the former behaviours may often be manifested in, or compensated by the motivations favoured in the latter. Correspondingly, behaviours which directly or indirectly benefit the individual donor may or may not be motivated by self-interest. Nevertheless, in evolutionary terms, any immediate benefit to the organism's *adaptive capacities* (rather than exclusively reproductive benefits) may qualify as being in the interests of its survival, given the essential phenotypic plasticity described in Chapter One. Such self-interest or self-concern need not be equated with the use of 'selfishness' to describe a lack of moral concern, unless that concern is evidently lacking where it should ordinarily be aroused i.e. in fulfilling the cognitive and affective conditions described above.

Given the different domains and levels of selection, the evolution of altruism need not demand such a degree of "fitness sacrificing behaviour" that could not be compensated for at another level or in another domain. Today, parents in many societies still regard children primarily as potential sources of income. To a lesser degree, a parent living in a middle-class suburban culture expects some reward for all the time and expense invested in bringing up a child. But the potential rewards need not be the proximate reason for such an investment. Children also satisfy the emotional needs of parents. It does not matter that natural selection has equipped us with genes that urge us to reproduce. Therefore, a genuine concern for the welfare of one's kin need not be considered as driven by some ulterior 'genetic' impetus. The emotions that motivate reproduction are still genuinely human, be they love or pure lust. But for this very reason, they are not moral sentiments. People do not have children as a moral duty. It is in our dealings with those who do *not* share our familial genes that morality has its purpose.

The plea to "love thy neighbour as thyself" is the best expression of the moral instinct for self-preservation, as it is based upon the understanding that human behaviour is inherently self-interested. Furthermore, it is a wise policy for a self-interested being to adopt because it promotes survival by co-operative means, which is demonstrably more effective than all-out competition. Living in large groups enables each individual to secure greater long-term benefits that could not be provided by the combined efforts of kin members. The additional resources available from other families and tribes made co-operation an essential aid to survival. But without the familial bonds of trust and care, dealing with strangers does not naturally incline toward co-operation. Groups descended from distinct lineages would normally live apart and compete for resources. In these conditions, how could any indiscriminate form of altruism have evolved? In short, it could not, which is why pure altruism is a rarity in any culture. What has survived instead is a conditional form of altruism that expects reciprocation i.e reciprocal altruism.[76]

In the past, a number of theorists, including Darwin himself,[77] have appealed to 'group' selection processes to account for the fitness benefits of altruistic acts that would otherwise seem to incur debilitating costs for the altruist. While no biologist can doubt that natural selection may produce

"significant cumulative change" to traits in whole populations, by the 1960s the theoretical workings of group selection were thoroughly discounted by the eminent biologist George Williams in his classic *Adaptation and Natural Selection*.[78] Thereafter, group selection theories were rightly discouraged, given Williams rigorous arguments and counterexamples showing that any purported 'group' benefit can more readily be accounted for at the level of selection upon genes and crucially, in contributing to the well-being of *individuals*. With regard to altruism in humans, Williams main argument is summarised in the following passage:[79]

> an individual who maximizes his friendships and minimizes his antagonisms will have an evolutionary advantage, and selection should favour those characters that promote the optimization of personal relationships…There is theoretically no limit to the extent and complexity of group-related behaviour that this factor could produce, and the immediate goal of such behaviour would always be the well-being of some other individual, often genetically unrelated. Ultimately, however, this would not be an adaptation for group benefit. It would be developed by the differential survival of individuals and would be designed for the perpetuation of the genes of the individual providing the benefit to another.

It should be noted that the validity of Williams' argument does not even depend upon accepting the latter claim that such a disposition should be genetically transmissible. As Darwin himself suggested, "love of praise and the dread of blame" may suffice to induce altruistic habits to the effect that a man "might thus do far more good to his tribe than by begetting offspring with a tendency to inherit his own high character."[80] However, contrary to Darwin's interpretation, this benefit to the tribe is more accurately described as the incidental *effects* of the *individual's* "love of praise and dread of blame" upon a significant number of other like-minded tribal members, rather than the expression of an homogenous group trait subject to competition with other homogenous groups.[81]

Sober and Wilson's recent revival of group selection is presented in terms of "multi-level" selection processes whereby altruistic behaviour at the group or tribal level can come to dominate selection for selfish behaviour at the individual level.[82] However, in order to defend the hypothesis of selection between groups, it is not enough to simply cite evidence for competition among tribal populations, or to assert that social norms are rigidly enforced or followed. Evidence for such selection would require that populations be largely homogenous in representing the beliefs, ideas and practices of individual members, which is simply inconsistent with the substantial evidence cited above, showing frequent intermingling and trade by which moral and political systems *evolve new rules and practices*. Despite acknowledging that "most early ethnographies…overemphasise the importance of social norms in tribal societies. Further research on the same societies often reveals more flexible and individualistic aspects of behavior", Sober and Wilson nevertheless proceed to cite precisely this dubious evidence to support *contradictory* assertions such as "Individuals are not free agents, however, but are so regulated by social norms within each lineage that they often act more like organs than organisms."[83] Apart from this astonishing contradiction, this deterministic depiction of agents as organs of cultural traditions is fundamentally inconsistent with the phenotypic plasticity on which human agents depend in order to conceive *new* ideas and beliefs and to *learn* from past experiences. In his critique of the depictions of

early ethnographers, Fried draws attention to the "tremendous gap" between the tribal ideals or "reinforcing mythology" and the "social reality" indicating that sanctions are often not enforced and norms irregularly followed.[84]

Quite apart from the ecological problem of setting rigid boundaries to the stable formation of groups and in determining exclusively group-defined behaviours,[85] Sober and Wilson conveniently discount the evolutionary benefits to the individual altruist and the variation among individual contributions to group benefits.[86] Furthermore, the multi-level approach does not in the least undermine the standard individualist account, as both explanatory models ultimately yield the same predictions. Yet, only the individualist model can adequately capture the level of *strategic* interaction among *individuals* with *competing interests*, as occurs in the Prisoner's Dilemma scenario,[87] which, ironically, Sober and Wilson cite as evidence for group selection.[88] The late John Maynard Smith raises the same objection noting that "the fitness of a group is merely the arithmetical sum of the fitnesses of the individuals that compose it".[89] To their credit, Sober and Wilson now seem to have largely conceded much of this criticism, insisting that in *Unto Others* "we do not claim that individual-level statements about the evolution of altruism are always unexplanatory and always fail to capture causal relationships", arguing instead for a "pluralism of perspectives" on the question.[90] Indeed, this is consistent with their earlier admission that "Human groups do not invariably function as adaptive units, and human nature may be as well-suited to maximising an individual's relative fitness within groups as it is to forming a group-level organism."[91]

For much of human history, encounters between members of different kin groups have been characterised by competition and conflict, even open warfare.[92] For the earliest human groups, the sheer scarcity of resources would have made co-operation highly dangerous, if not suicidal. Under such conditions, outsiders can only ever be seen as a threat, which indeed they were. Even today, many tribes conduct raids against their neighbours in a ritual fashion.[93] But unlike their forebears, these modern tribesmen have also learnt the benefits of co-operation in exchanging valuable resources. At some point in their evolutionary history their ancestors, like ours, began to grow crops and domesticate animals, which afforded them much greater control over the problem of scarcity. As a consequence, more and more children survive into adulthood and over several generations the population will increase and eventually stabilise to a level of subsistence.[94] If the environment is particularly fertile or abundant in resources, then the tribe has every incentive to exploit it for all it is worth. Armed with an overabundance of goods, or simply an oversupply of one staple crop or animal, this tribe need no longer fear outsiders. For every member of such a community, it is in their interest and their power to form an economic alliance with other communities. As long as one group has something another lacks, there will always be a sound economic basis for co-operation.[95] A trading partnership will be of mutual benefit to members of both parties. As long as there is a balance of power, pure competition is not an option, as it can only result in a long war of attrition, with great risk to the survival of either side. Thus, evolution favours strategic co-operation at least as much as competition.[96]

In the same way that natural selection determines the winners competing in the game of life, it also determines the game's rules of engagement. Co-operation pays on some occasions, less so on others. Sometimes it may have a neutral effect, at other times it will be suicidal. Encounters between

members of different kin groups would have been dangerously unpredictable, so the process of forging alliances could not have evolved by purely rational deliberation. There is no need to imagine a prehistoric social contract being somehow negotiated and handed down like the Ten Commandments. Instead of appealing to unlikely myths to explain the evolution of co-operation, we should look to the very principle that produces evolution generally - natural selection.

The great mathematician Von Neumann developed 'game theory' as a means of determining the probable outcomes of competitive interactions between rational players.[97] The heuristics by which players make their moves in a game of chess, for instance, are informed by calculated estimates of the opponent's available options. If each player were equally proficient and accurate in predicting the other's next move, then a stalemate will eventuate. But of course, in the game of chess, as in the game of life, encounters do not begin as evenly matched affairs between like-minded players. Natural selection ensures that there is sufficient variation in the skills of the contestants, even more so in those from divergent populations. Just as individuals vary in size, acuity of senses, agility etc., they also vary in intelligence and temperament.

Most of us have learnt from bitter experience that it does not pay to be too trusting of strangers. Yet co-operation can never evolve unless one party makes an overtly trusting gesture toward a stranger. Thus, a relationship based on reciprocal altruism must begin with an element of risk for the party who first signals their willingness to benefit the other's interests. For there can be no guarantee that the other will even accept the offer, still less that it will ever be reciprocated. This would not have been such a problem for neighbouring tribes or clans, as their populations are small enough to keep track of the fluctuating debts and credits that accrue among different individuals. Anyone who fails to repay some benefit received is soon identified as a cheat and punished.[98]

As co-operative contacts among tribes grow closer, members intermingle to the point where it becomes mutually advantageous to live in a single large community. As long as there are sufficient local resources available to sustain a growing population, then it becomes possible to use a combined labour force to greatly reduce the time and energy devoted to harvesting those resources. In such a village community, social mores evolve directly in relation to the benefits contributed by each member to every other member. As a rule, the most valued members are the most useful in meeting the needs of others. Therefore, genuine altruists i.e. those who do not expect any reward for their efforts should thrive in a village where everyone knows everyone else. Why then are they so rare?

As John Maynard-Smith noted, in any sizeable population of mixed descent, natural selection guarantees that there will be enough variation to contain a wide spectrum of individual temperaments. There will be some Doves, who are invariably co-operative and some Hawks who take but never give. However, the bulk of the population will exhibit more flexible behaviours between these two extremes. In the long run, it pays less to an inflexible Hawk or Dove than it does to adjust your behaviour to the demands of each particular encounter.[99] The completely selfless altruist who is dedicated to serving the welfare of others inevitably suffers exploitation and exhaustion at the hands of the ruthlessly selfish. Hawks will soon proliferate at the expense of Doves. But once the Doves become extinct the Hawks are deprived of their easy prey, so they too become victims of their own inflexible behaviour. As shown in the discussion of phenotypic plasticity in Chapter One, such inflexibility is an evolutionary

liability for most species given the inevitable disruptions to their habitats. As a result, it is not surprising that the relatively rare 'ultra-social' species i.e. ants, bees, wasps, termites are also those whose members are often well-equipped to insulate themselves from the harshest ecological conditions[100] and produce the most genetically similar offspring, usually from a single queen. As such, they have the *least* need of phenotypic plasticity (as exemplified by the four simple rules governing the behaviour of foraging ants), and in functioning as a superorganism all the members benefit from the reproductive altruism supporting the colony, with such close inbreeding ensuring that their social behaviour is correspondingly under the *most* genetic control instead.[101]

Evolutionary game theory unequivocally demonstrates that a pure strategy of selfishness is ultimately counterproductive, for the precise reason that it exhausts the goodwill of others. As dedicated altruists gain nothing in their dealings with purely selfish persons, they cannot survive without some support from other like-minded contacts. Thus the altruist cannot afford to refuse at least some benefits from the altruism of others. By necessity, evolution will favour the kind of altruists who are equally willing to give and receive. Unlike the pure altruist, they will not be exploited for their generosity but will themselves benefit from it being returned by other like-minded individuals. As Maynard Smith notes, evolutionary game theory is "based upon a well-defined dynamics – the evolution of the population – and the 'solutions' of the game are the stable stationary points of the dynamics."[102] Accordingly, it is particularly well-suited to modelling the evolution of stable rule-governed behaviours in human societies where strategic rationality is crucial to maintaining co-operation. This is exemplified in Maynard Smith's discussion of the normative stability shown in the act of queuing, wherein individuals benefit from each other's reciprocal willingness to punish queue-jumpers.[103] Of course, the strength and reliability of the norm will tend to vary only within certain parameters that are already partially determined by each individual's initial propensity for following the rule to queue on the appropriate occasions. However, far from undermining the validity of such models, evolutionary game theory succeeds by showing how different strategies are more or less likely to evolve within various social ecologies. Therefore, the fact that a simple rule like Tit-for-Tat can succeed against such a diverse range and proportion of mixed strategies clearly demonstrates its robust nature (TFT is discussed directly in the next chapter in connection with more recent models evaluating its performance in a more complex social environment).

The strategy of reciprocal altruism greatly reduces the chances of being exploited by selfishness. For the reciprocal altruist will soon cease contact with those who repeatedly fail to return favours. Of course, some people will be more forgiving and indulgent than others, but everyone has their limits. Even the closest relationships cannot survive the need for reciprocity. Marriages and friendships routinely collapse when one partner's investment gains little reward from the other. In this context, the expression "what are friends for?" is not a cynical remark but a wise reminder of the true value of mutual aid - that it is more beneficial for all of us to help each other than to help ourselves.

Recalling an earlier point, it is important to note that the evolution of reciprocal altruism in human populations would be *expected* to be supported and stabilised by robust mechanisms selected at the psychological level, in addition to lower level genic forces. Accordingly, in formulating his theory, Trivers took a scrupulous approach in testing its functioning at the motivational level by comparing its

psychological and behavioural predictions with evidence accrued from psychological, sociological and anthropological data.[104] The results indeed confirmed that motivational factors such as "friendship and the emotions of liking and disliking…moralistic aggression…gratitude, sympathy and the cost/benefit ratio of an altruistic act…guilt and reparative altruism…detection of the subtle cheater: trustworthiness, trust and suspicion" are all in keeping with the theory's predictions.[105] Furthermore, reciprocal altruism is *not* a theory that invariably reduces psychological altruism to selfishness in evolutionary terms. Trivers explicitly cites its consistency with the anthropological data indicating that humans "tend to respond more altruistically when they perceive the other as acting "genuinely" altruistic, that is, voluntarily dispatching an altruistic act as an end in itself, without being directed toward gain" and that "help is more likely to be reciprocated when it is perceived as voluntary and intentional".[106] As such, it is perfectly described as a conditional form of altruism in humans, as it is only likely to evolve on such motivational conditions.

The practice of reciprocal altruism evolved in human societies as an adaptive response to the demands of communal life. By adopting a flexible, lifetime strategy of mutual aid, most members will incur fewer costs and reap greater rewards than by any other general approach. As an historical development that has been reliably modelled by evolutionary game theory, reciprocal altruism can be seen as the basic principle governing all prevailing moral codes. In spite of the apparent diversity of moral prescriptions and prohibitions, all human societies function within the economic boundaries inherited from their ancestors' adaptive success. As the most economically stable strategy for humans living in extended groups, the practice of reciprocal altruism would be a necessary adaptation for all such groups.[107]

Anthropologists have long recognized that the system of moral beliefs and practices in a society is invariably linked to its system of work and resource distribution. Despite some rare and short-lived exceptions, the moral norms function as prescriptions for reproducing those behaviours that best serve the established economic system. The myths, rites and taboos that differ across cultures merely reflect different traditions of vested interests. Far from providing an argument against the genetic origin of moral practices, these traditions show their self-directed structure only too well, as they generally yield some benefits for those who abide by them. As long as most individuals continue to benefit more by reproducing inherited moral practices than they would by defecting, there will be no significant moral progress within such a society. This is why surviving tribal societies still function by strict adherence to ancient beliefs and traditions that have stood the test of time. However, as argued earlier, the persistence of such ancient traditions need not be cited as evidence for group selection processes, when their most stringent moral rules are all devoted to the principle of self-preservation: hence the classic taboos against incest and eating foods that have been "tainted" by enemy tribes.[108] Such rules survive because those *individuals* who obeyed them most often survived in greater numbers than those who did not. Many of these practices may well turn out to be founded on superstition. Nevertheless, as long they contribute to the well-being of those who observe them, they will inevitably grow ever deeper to the point where they represent the unquestionable "sacred" commandments of the ancestors.[109]

While different cultures express different sets of moral imperatives, those differences are largely superficial variations on one major theme: conducting co-operative relations between individuals with competing interests. Abstracting from the local and historical conditions that produce variations in moral practices reveals a striking conformity in their function. Virtues and vices are typically reducible to their utility in producing or preventing co-operative self-interest. In this light, consider the four "natural" virtues described by Plato: Justice, Fortitude, Prudence and Temperance.[110] Whether or not these encompass the definitive range of virtuous acts is, of course, open to question. But they are all universally understood as qualities that are morally desirable. Nowhere is there a culture that finds no particular merit in promoting the principle of justice. As justice is primarily concerned with balancing debts and credits between different parties, any society that requires co-operative relations between non-kin members is bound to conceive just acts as a necessary virtue. Precisely what constitutes a just act may differ between cultures but this should not be taken as an indication that justice itself is weakened by relativism. Historically, justice has primarily functioned to ensure a level of social harmony sufficient to maintain a society's economic productivity.[111]

Economic necessities also make temperance a virtue for every society. Even those with the most abundant resources have learnt the value of moderating consumption to allow for any environmental change that might suddenly reduce supplies. History shows that the greatest empires are likely to have collapsed as a result of the intemperent exploitation of their environment. Both the Mayan and Roman civilisations bear witness to a failure that is at once economic and moral.[112] These tribes could not have evolved into huge populations except by constantly taking possession of an ever increasing supply of land and materials. Here the virtue of fortitude has been overdeveloped so that its initial value in defending tribal interests has been put to greater use in appropriating the resources of foreign tribes. Yet again, necessity appears as both the material and psychological factor that accounts for the different emphasis in the value of fortitude. As Nietzsche recognised, the qualities that are best suited for winning wars can only be developed and reinforced by cultivating the Homeric virtues that equate nobility with strength of mind and body.[113] Thus, it is no accident that an empire culture should emphasise the value of fortitude at the expense of temperance. While both traits contribute to the welfare of most citizens, great fortitude is the more crucial for the maintenance of an empire. As such, the qualities of strong leadership are taught as the 'highest' virtues, while those that are more common 'herd instincts' are of lesser value.[114]

Prudence is a trait instinctive to most animals. As a mechanism for avoiding danger, it has an elementary fitness value. More than any other trait, being prudent "makes a virtue of necessity" given that survival generally depends upon it. But while Plato recognised prudence, justice, temperance and fortitude as natural virtues, his conception of nature incorporated virtues as expressions of the ideal 'form' of the good.[115] Plato's theory of forms gives virtues a natural form only on the belief that they reflect a higher form of goodness that is beyond human understanding. This view that moral concepts have a transcendent reality has had considerable appeal in philosophy, religion and 'folk psychology'. It is hardly surprising that such a faith prevails in mythologies and religions where the need for purpose and meaning gives prejudice to the search for moral 'truth'. Philosophers, however, have no excuse for reverentially accepting received ideas on any subject. On the contrary, it is their self-appointed task to

subject standard beliefs to severe scrutiny. Dogmatism should be anathema to the philosophical mind, yet throughout history, otherwise great philosophers have been seduced by the beauty of an idea which comforts and confirms their own preconceptions.[116] In opposition to this tradition, the next chapter will show how the foregoing account of the evolutionary function of moral and political systems presupposes a universally teleological process of pragmatic moral evaluation that is necessarily prescriptive and consequentialist in its concern with promoting the common prudential interests of agents in any society.

[37] Lionel Tiger & Robin Fox, *The Imperial Animal* (New York, Henry Holt, 1989), pp.124-125

[38] Immanuel Kant, *Critique of Practical Reason*, trans. L. W. Beck (New York;
Toronto, Macmillan, 1993), p.20

[39] See De Waal, *Good Natured*, pp.128-132 & 150ff. & De Waal's earlier *Chimpanzee Politics: Power and Sex among Apes* (Baltimore & London, Johns Hopkins University Press, 1989), pp.205-207 & pp.210-213; See also Margaret Power, *The Egalitarians – Human and Chimpanzee*: *An Anthropological View of Social Organization* (Cambridge, Cambridge University Press, 1991), pp.4-9 outlining commonalities in human and chimpanzee social systems

[40] See Matt Ridley, *The Origins of Virtue*. (London, Viking, 1996), pp.106 & 108; Tim Megarry, *Society in Prehistory: The Origins of Human Culture* (London, Macmillan, 1995), pp.276-277; Stephen Mithen, *The Prehistory of the Mind*: *A Search for the Origins of Art, Religion and Science* (London, Thames & Hudson, 1996), pp. 169-171, 182-183

[41] Megarry, op. cit., pp.276-278; Mithen, op. cit., pp.173-175

[42] Megarry, op. cit., pp.277, 285 & 287

[43] Ridley, op. cit., pp.109-117

[44] Stephen Boyden,. *Western Civilization in Biological Perspective: Patterns in Biohistory* (Oxford, Clarendon Press, 1987), pp.113 & 127-8

[45] Mithen, op.cit., pp.173-175

[46] See recent article by Kate Douglas, "Born to Trade," *New Scientist*, Vol. 183, No. 2465, 18 September 2004, pp.24-28

[47] Christine A. Hastorf, "One Path to the Heights", in *The Evolution of Political Systems*: *Sociopolitics in Small-Scale Sedentary Societies*, ed. S. Upham (Cambridge, Cambridge University Press, 1990), pp.159ff.

[48] Ibid. pp.161-162

[49] Ibid.

[50] Ibid., pp.163-168

[51] Ibid., p.168

[52] Megarry, op. cit., pp.216-217

[53] Power, op. cit., p.18

[54] Morton H. Fried, *The Evolution of Political Society*: *An Essay in Political Anthropology* (New York, Random House, 1967), pp.88-89

[55] Bruce G. Trigger, "Maintaining Economic Equality," in *The Evolution of Political Systems*: *Sociopolitics in Small-Scale Sedentary Societies*, ed. S. Upham (Cambridge, Cambridge University

Press, 1990), p.137.

[56] Christopher Boehm, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Cambridge, MA; London, Harvard University Press, 1999), pp.73-74

[57] Ibid., pp.60-61; see also De Waal, *Good Natured*, pp.137-8; cf. Wright, op. cit., p.71-72 for opposing interpretation of egalitarianism as practised by the !Kung

[58] Colin Turnbull, "Demography of Small Scale Societies," in *The Structure of Human Populations*, eds. G. A. Harrison & A. J. Boyce (Oxford, Clarendon Press, 1972), p.31, cited in Power, op. cit., p.20

[59] Boyden, op. cit., pp.133-134

[60] J. L. Mackie, *Ethics: Inventing Right and Wrong* (Harmondsworth; New York, Penguin, 1977), pp.21 & 24; 35

[61] G. E. Moore, *Principia Ethica*, rev. ed. (Cambridge, Cambridge University Press, 1993), pp.50-58

[62] David Hume, *A Treatise of Human Nature*. ed. L. A. Selby-Bigge (Oxford, Clarendon Press, 1975), I. III. VI; III. I. I, pp.465-466

[63] Thomas Nagel, *The View from Nowhere* (New York, Oxford University Press, 1986)

[64] R. M. Hare, *The Language of Morals* (Oxford, Oxford University Press, 1964), pp.2-3

[65] More precisely, it is an adaptation of an adaptation - what Gould & Vrba call an "exaptation" – See Dennett, op. cit., p.275

[66] See Dennett, op. cit., pp.379-380

[67] C. Daniel Batson, *The Altruism Question: Toward a Social-Psychological Answer*, (Hillsdale, NJ, Lawrence Earlbaum Associates, 1991) p.75

[68] Ibid.

[69] Ibid., p.83

[70] W. H. Durham, *Coevolution: Genes, Culture, and Human Diversity* (Stanford, Stanford University Press, 1991) p.16

[71] Richard Dawkins, *The Selfish Gene* (London, Paladin Granada, 1976), pp.101-2

[72] Ibid., pp.4-5

[73] Durham, op. cit., pp.13-14 & pp.82-85 for case studies

[74] Michael Argyle, *Cooperation: The Basis of Sociability* (London, Routledge, 1991), pp.143-5

[75] Dennett, op. cit., p.478

[76] See Robert Trivers, "The Evolution of Reciprocal Altruism," in *Quarterly Review of Biology*, vol. 4, pp.35-57

[77] Charles Darwin, *The Descent of Man* (2nd ed. Amherst, NY; Prometheus Books, 1998; original ed. New York, Cromwell, 1874), p.137

[78] George C. Williams, *Adaptation and Natural Selection: A Critique of Some Evolutionary Thought* (Princeton, NJ, Princeton University Press, 1966), p.109; see also William's more recent *Natural Selection: Domains, Levels and Challenges* (Oxford, Oxford University Press, 1992), pp.45ff

[79] Williams, *Adaptation and Natural Selection,* p.94

[80] Darwin, *The Descent of Man*, p.136

[81] Ibid., p.137

[82] Elliott Sober & David Sloan Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior*, (Cambridge, Mass., Harvard University Press, 1998), pp.32-33

[83] Ibid., pp.163 & 174;

[84] Fried, op. cit., pp.145-146 & 148

[85] Ibid., pp.154-174, esp.p.168 criticising the very conception of a distinct 'tribe', given the fluid interactions with other 'tribes', migration etc.,

[86] Benjamin Kerr & Peter Godfrey-Smith, "Individualist and Multi-level Perspectives on Selection in Structured Populations," in *Biology and Philosophy* 17, 2002, pp.480-481 & 485

[87] Ibid., pp.479, 494-5

[88] Sober & Wilson, *Unto Others*, p.152

[89] John Maynard Smith, "Commentary on Kerr and Godfrey-Smith," in *Biology and Philosophy* 17, 2002, p.526

[90] Sober & Wilson, "Perspectives and Parameterizations: Commentary on Benjamin Kerr and Peter Godfrey-Smith's "Individualist and Multi-level Perspectives on Selection in Structured Populations," in *Biology and Philosophy* 17, 2002, p.529

[91] Sober & Wilson, *Unto Others*, p.192

[92] Boyden, op. cit., p.132

[93] Tiger & Fox, op. cit., pp.213-214; See also Durham, op. cit., pp.375-379 for case study

[94] Boyden, op. cit., pp.115&118

[95] Ibid., pp.129-30

[96] Ridley, op.cit., pp.199-201

[97] John Von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton, Princeton University Press, 1964)

[98] De Waal, *Good Natured*, pp.159-61; See esp. Gerd Gigerenzer, *Adaptive Thinking: Rationality in the Real World* (Oxford, Oxford University Press, 2000), p.217ff.

[99] John Maynard Smith, *Did Darwin Get It Right?: Essays on Games, Sex and Evolution* (London, Penguin, 1993), pp.194-5 & 206-15

[100] e.g. as termites insulate themselves by creating "their own stable microclimate within their mounds" – see Kim Sterelny, *Thought in a Hostile World: The Evolution of Human Cognition* (Malden MA; Balckwell, 2003), p.149

[101] See Robert Trivers, *Social Evolution*, (Menlo Park, CA; Benjamin/Cummings, 1985), pp.169ff

[102] Maynard Smith, *Did Darwin Get It Right?*, p.195

[103] Ibid., pp.198-199

[104] Trivers, *The Evolution of Reciprocal Altruism,* pp.47-54

[105] Ibid., pp.48-50

[106] Ibid., p.51

[107] Richard Alexander, "A Biological Interpretation of Moral Systems," in *Issues in Evolutionary Ethics*, ed. P. Thompson, (Albany, SUNY Press, 1995), pp.188-191

[108] Steven Pinker, *How the Mind Works* (New York, Norton, 1997), pp.385 & 456; See Durham, op. cit., Ch.6, for contemporary analysis

[109] Edward O. Wilson *On Human Nature* (Cambridge, Mass. ; London, Harvard University Press, 1978), pp.183-1844

[110] Plato, *The Republic*. ed. H. D. P. Lee (Harmondsworth, Penguin, 1974) 428

[111] Robert Wright, *Nonzero : The Logic of Human Destiny* (London, Little Brown, 2000), pp.99-179

[112] Ibid., p.154; Boyden, op. cit., p.100

[113] Friedrich Nietzsche, *On the Genealogy of Morals*, in *Basic Writings of Nietzsche*, trans. & ed. W. A. Kaufmann (New York, Modern Library, 1992), pp.469 & 476-478

[114] Ibid., p.462

[115] Plato, op. cit., 472-483

[116] Even Kant's categorical imperative can be considered as an ascetically, if not aesthetically appealing example of such transcendental moral foundationalism

CHAPTER THREE

PART TWO:

THE PRAGMATISM OF MORAL AGENCY

CHAPTER THREE

THE TELEOLOGICAL NATURE OF MORAL VALUE

(i)   Prescriptivity and Consequentialism

As moral beliefs are often held as deep convictions that carry a strong emotional attachment, they are particularly averse to careful analysis. No-one wants to discover that a long cherished belief is groundless or even weak, especially if it has had a central prescriptive force in the conduct of their lives. But giving moral categories a transcendental quality is philosophically inadequate, inasmuch as it gives the philosopher an excuse to avoid any empirical examination of moral phenomena that may show it to have humbler origins. Accordingly, the meta-ethical evidence cited in the previous chapter supports a biological functionalism in moral judgement and practice which limits both the intelligibility and practicality of normative moral theories which attempt to discount or transcend such basic and indispensable functions. Therefore, this chapter will consider the extent to which certain inherited principles, norms and virtues are indispensable in serving the common interests of autonomous agents in any viable society, thereby challenging the practical value of alternative moral and political systems which would discount such basic values.

In our daily lives, most of us cannot afford to be moral skeptics. If only to avoid harm, we each need to able to recognise harm as such. Against the obscurantism of Plato's metaphysical interpretation of morality, we have Aristotle to thank for putting the subject into a broader perspective that takes account of the biological constitution of human nature. More than two thousand years before Darwin, Aristotle noticed that living things were essentially characterised by their apt design for performing specialised tasks.[117] In the absence of a scientific method or ethos, Aristotle was unable to discover any causal principle to explain natural adaptation. Therefore, he was forced to conclude that species were designed for a set of purposes in advance. With the principle of natural selection, biologists can now provide a causative account to replace the teleological argument. Even so, Aristotle was certainly right in his understanding that species' behaviour is essentially purposive. Autonomous behaviour can only be comprehended teleologically, even if strictly speaking an organism's future goals cannot be the cause of its genetic constitution. In recognising that life is a process of striving to achieve a state of "well-being", Aristotle placed the study of ethics back in the real world, where virtues have a practical purpose.

Over the following centuries the study of ethics was increasingly appropriated into the realm of metaphysics and theology. Good and evil remained absolute truths that could not be grasped by any empirical study of human nature, but by revelation from God. Only towards the end of the Enlightenment did ethics begin to be seen as a legitimate subject for empirical investigation. By the time Hume published his moral theory philosophers no longer needed to appease the Church with specious supplementary arguments for God's moral authority. At the same time, many philosophers on the continent remained loyal to the Platonic tradition. To this day, the opposing views of Hume and Kant continue to dominate ethical theory with a rediscovered Aristotle filling the breach with modern 'virtue ethics'. The proponents of the latter theory have resurrected the principle of virtue to support an agent-centred approach to guide moral deliberation. But they have conveniently neglected Aristotle's interest in the biological process that makes virtue a necessity for the achievement of well-being.

For convenience, ethical theories are typically characterised in terms of their conformity to the Humean or Kantian legacy. As long as it is agreed that reason is a prerequisite for moral decision-making, then any plausible moral theory is basically definable by the role that it assigns to reason. For Hume, reason is the "slave of the passions" which leads to a conclusion that moral judgments are disguised sentiments.[118] For Kant, the moral law is a product of reason designed to govern those instinctive passions.[119] Thus the two are poles apart but they both agree that moral rules are necessary for the realization of human ends. Kant's categorical imperative assigns the highest moral value to the agent's willing performance of moral 'duties' without regard to consequences. Hence Kantian ethics are usually described in opposition to consequentialist varieties. But in formulating the criteria by which practical reason determines these duties, Kant is forced to concede that consequences of necessity bear directly upon judging which actions are dutiful.

How is it possible that the moral law can determine which actions must be performed yet remain ignorant of their consequences? Such a law could scarcely be moral when it shows no concern for its effects. Still less is it practicable to act without any interest in the outcome. Even if it were possible to become blind to the consequences of one's actions, would it be at all desirable? The blind

obedience required by the moral law turns out to be contrary to practical reason, the very opposite of Kant's claim. Ultimately Kant undermines his insistence upon the categorical requirement of the moral law when he produces a set of maxims whose prescriptive force is derived solely from the demands of practical reason. Reason affords a test for the supposed viability of any moral action. Thus lying is deemed to fail this test on the grounds of logical inconsistency. If the habit of lying were universally condoned, it would no longer be possible to trust anyone's word. Everyone would suffer the consequences if such a policy should become a universal law.[120] By his own example then, Kant demonstrates that it is, after all, the undesirable *consequences* foreseen that disqualifies lying as universal practice. However, there is no *logical* contradiction that prevents the practice of lying by a single individual. Neither the occasional nor the inveterate liar is interested in convincing others to follow suit. But this is not because of some failure to grasp what would happen if everyone did so. On the contrary, given that lying is commonly aimed at securing an advantage over others, the habitual liar has every reason to hope that others do *not* follow the same policy to the same degree.

Without due consideration of consequences, reason has no means and no prospect of determining which actions ought to be pursued under any circumstances. Any conception of moral value as absolute does not withstand the scrutiny of empirical investigation into the actual conditions that generate moral questions. Moral value cannot be distilled and quantified into a set of formal rules to be applied by all persons, at all times, in all places. It is precisely these hypothetical variables that make moral judgments so weighty, so *consequential*. Once consequences are subtracted from questions of moral value, there is no value left to judge. With clear duties to be applied in all circumstances, regardless of the consequences, moral choice is dispensed with. Moral value is set in stone, its hypothetical character erased and moral action requires no more and no less than obedience to immutable laws guaranteed by the 'authority' of religion or tradition.

The evolutionary reconstruction of moral value as a development of practical necessity gives full recognition to the imperative weight of moral claims, while preserving their hypothetical import. Despite Kant's misgivings, it is not a tragic loss for humanity to admit that all moral claims are only hypothetical imperatives. There is nothing to be gained from insisting that moral directives must apply either categorically or absolutely, when the human mind has evolved to reason hypothetically in the light of individual goals. In fact, adherence to an absolute, transcendental conception of morality not only misrepresents its practical value in meeting human needs, but a system of categorical moral beliefs is positively dangerous and *immoral* to the extent that it encourages a view of moral behaviour that denies human freedom. History has shown that an absolutist conception of morality is liable to produce an equally absolute moral fervour in imposing that conception on outsiders and dissidents.[121]

This sense of moral certainty that invests categorical moral claims is, admittedly, a common feature of moral thinking in most societies. While our experience of moral dilemmas reminds us of the hypothetical interests that compete for our attention, we are often enough quite sure of the right action to be taken. Furthermore, when all members of the group express a consensus on the rightness of an action, does it nevertheless remain only a hypothetical imperative? If everyone in the group agrees that lying is wrong in all circumstances, then, as far as they are concerned, it would appear to express a categorical imperative.

While some anthropologists might be satisfied to describe a society's normative moral beliefs in categorical terms of acceptance, such a naïve depiction can hardly represent the actual degrees of conviction expressed by each individual. In any case, those moral beliefs that are held with absolute conviction by all members of one group do not have any binding force on members of other groups who would dispute such claims. Only the most fanatical religious adherents demand that every human being is obliged to obey their commandments. Most cultures and religions in fact define themselves on the basis of the exclusivity of their moral claims.[122] For instance, a Jew does not forbid a Christian to eat pork. The Jew understands the limited nature of the injunction, thereby acknowledging its hypothetical application.

This relativity of moral judgments undermines any attempt to describe them as categorically i.e. unconditionally binding. For they can only ever be binding if and when they are assented to. In this respect, freedom of the will guarantees that all moral judgments are hypothetical, even those that appear to demand universal assent. As all such judgements depend upon the agent's ability to decide *if* a belief or action should be endorsed or acted upon in a given set of circumstances, no *particular* moral judgement can be categorically binding.[123] As Philippa Foot notes, an essential feature in the constitution of a moral judgement or action is the exclusion of any physical or psychological compulsion.[124] After all, even in Kantian terms, it is the agent's *autonomy* in transcending such compulsions that warrants categorical obedience to the moral law. Nevertheless, that obedience itself remains possible only on the *hypothetical* condition that the agent should *willingly intend* to follow any particular moral directive. This is well illustrated in Foot's critique, where even the rules of etiquette would qualify as categorical imperatives, given that their observance is required regardless of the agent's own personal inclinations.[125] A contemporary illustration can be witnessed in the evolving norms governing personal behaviour in a public setting such as a restaurant or cinema. A sign requesting persons to refrain from smoking or to avoid using a mobile phone is not premised on any proviso that a person should *desire* to act accordingly. However, it *is* premised on the assumption that a person be able to control their inclinations and therefore also presupposes that persons must be able to *choose* to follow such directives. Hence, as Foot suggests, the description of any such imperative as being "inescapable" or having some "binding force" is simply incoherent, as their observance invariably depends upon the respondent's *own judgement* in choosing whether or not to follow them on any particular occasion.[126] In this respect, moral judgements are no less dependent upon the agent's autonomous decision to respect a particular moral rule or principle, as implied in the notion of moral responsibility.

Further to this point, if moral imperatives were truly inescapable, no person would be capable of judging that they may be ignored on certain occasions, which is obviously not the case. But again, this only serves to indicate that any given rule can only be binding *if* an individual chooses to accept it as a constraint upon their current or future behaviour. Thus, quite apart from the prudential risks involved in following rules that discount the agent's own powers of judgement, the concept of the categorical imperative is refuted by the commonplace psychological and behavioural evidence against its reputed 'binding force'. Although it may be objected that persons often enough *feel* psychologically bound to always act in accordance with some moral rule, that person's closest neighbours who have

learnt the very same rules in the same culture, may nevertheless apply them more judiciously, more or less aware that the 'force' of such imperatives is determined by their own personal will. In any case, as Foot rightly points out, the recurrent psychological feeling of being bound by a rule or norm can hardly be cited as evidence for any categoricity inherent in the rule itself, particularly when such feelings may be easily attributed to traditional methods of moral education.[127] Similarly, the 'compulsive' gambler who reports feeling that they categorically "must" place another bet can hardly be cited as psychologically representative of the majority of occasional gamblers who do *not* exhibit such a lack of control over their immediate desires.

Not surprisingly however, many people continue to find this situation completely unacceptable. Like Kant, they fear that a morality that is explicitly hypothetical can have little value, as it amounts to nothing more than a system of engaging the most appropriate means of satisfying our desires without restraint. Fortunately, this is an unwarranted conclusion that is clearly refuted by everyday experience and observation. As social animals, human beings depend upon each other for survival. Restraint and co-operation evolved as a practical necessity. As the Hawks and Doves scenario showed, wanton selfishness soon becomes a liability for life within a mutually dependent group, as others become less and less accommodating. Of course, the prevalence of mechanisms favouring mutual dependence within a society provides no guarantee that dissidents or outsiders will not also be exploited or attacked. As noted earlier, there is no shortage of historical examples where loyalty and co-operation are marshalled for the most nefarious purposes.[128] Co-operation may often enough be used to *incite* conflict or prejudice, as in the case of the Nazi regime, or to exploit the trust of others, as in the case of an amoralist. Nevertheless, while co-operative virtues and social institutions cannot themselves be relied upon to generate the kind of moral sympathy and concern needed to prevent such abuses, *without* such virtues and institutions those abuses could not be countered. The Nazi threat itself, the practice of slavery, and acts of individual amoralism are only defeated by *concerted* action guided by *common* moral principles and ideals. In Part Three, it will be shown how the latter principles and ideals can be legitimately and pragmatically employed against the former threats on the basis of universal principles of rational coherence. At this stage, the modest intention is merely to illustrate the indispensability of co-operative virtues and values in the mutual protection they provide *against* the abovementioned dangers. As such, in considering the pragmatic value of co-operation, it is vital not to "throw the baby out with the bath-water", as it were.

The human capacity for autonomy is what makes moral imperatives hypothetical. Far from being a tragedy, this is surely cause for celebration. Imagine what life would be like if moral judgments did in fact possess this mysterious power to command. Moral acts would no longer be chosen but caused like an instinctive response to a stimulus. Hear the moral law and immediately obey. This is hardly the behaviour of highly rational, autonomous agents. If this is the highest expression of moral development, then moral systems would be akin to the sets of algorithms perfected by insects billions of years ago. Ironically, the autonomy that Kant claimed to be the means of willing obedience to the moral law is dissolved in the process of its assimilation. Instead of freeing human behaviour from the tyranny of instincts, as Kant had hoped, this total commitment of autonomy produces a moral vacuum. Fortunately, as long as we have autonomy, moral demands can only be hypothetical.

Nevertheless, though hypothetical, such demands remain, by definition, imperatives in the sense that *if* a person has willingly decided to follow a particular directive *then* coherence demands that it be acted upon insofar as circumstances permit such action. Thus, in deciding to take a drink, a thirsty person acts on the hypothetical understanding that the drink will slake their thirst to some degree. If that expectation is fulfilled, then arguably the person's autonomy is preserved or even enhanced in being 'bound' by this hypothetical imperative. But this should not be taken to contradict the claim that adherence to a *categorical* imperative threatens to diminish one's autonomy. The agent's autonomy consists in their capacity to exercise control in judging the value of possible actions, moral or non-moral, by their variable effects in distinct situations. Providing the agent has given due attention to those effects, acting in accordance with a hypothetical imperative allows the agent to *learn* from those experiences and to *adjust* their behaviour to conditions which might otherwise prevent them from realising their goals. By contrast, as a categorical imperative demands adherence *regardless* of the agent's own interests, it takes no account of consequences which may well harm those interests in certain circumstances.

With reference to the above example, following a categorical imperative to drink i.e. regardless of one's interests, moral or otherwise, is liable to cause irreparable harm to oneself and others. Like the hunger striker, if a person has resolved to minimise their consumption of water in the hope of achieving some moral end, that *hypothetical* ideal will certainly be compromised if that person regards the satisfaction of their thirst as categorically required whenever it is experienced. For large numbers of people inhabiting regions where potable water supplies are infrequently available, survival depends upon actively and regularly *suppressing* the urge to drink one's fill so that supplies are rationed for future consumption. So too, like the compulsive gambler, a recovering alcoholic is occasionally liable to believe that he categorically 'must' have a drink even while knowing that it is *not* in his best interests to do so. Although his immediate physical desire may be relieved, he may well *prefer* to be *rid* of that desire, in which case he cannot afford to ignore the hypothetical situations which may tempt him to succumb to its influence.

In spite of these psychological and prudential arguments against conceiving prescriptive judgements as categorical imperatives in the Kantian sense, it remains both possible and plausible to defend such judgements from a categorical basis which does *not* discount the agent's own motivations or interests. Arguably, logical and epistemological constraints on the validity of such judgements do provide the invariable conditions which categoricity entails. Thus, in Chapter Five it will be argued that the principle of non-contradiction does indeed furnish the grounds for a categorical requirement to preserve one's autonomy and well-being in order to act as an autonomous agent. However, as even that categorical requirement can only be demanded on the *hypothetical* precondition that agents willingly continue to conceive and pursue their goals, its categoricity remains *conditionally* binding.

As suggested above, no doubt there is a strong psychological need to believe that at least some of our moral beliefs must be unquestionably justified. Unless we are convinced that certain kinds of acts are never permissible, or that others are always necessary, all moral decisions would be taken with extreme trepidation. The decision to swat a pesky fly at a barbecue would arouse as much angst as launching a nuclear attack. Clearly, it is the foreseeable consequences that determine the moral value of

both acts. Ending the fly's life is immediately judged as morally insignificant in comparison to your eating pleasure. For most people, their lack of hesitation in killing a fly signifies their moral certainty. But this is not because they are following a categorical imperative that makes it either a duty or a right to kill members of some species. If that were the case, we would aim to kill every fly that bothered us, on every occasion regardless of the circumstances.

The ease and certainty involved in many everyday moral judgments is liable to make them appear unconditional. But once we reflect upon the circumstances in which the choice is made, it becomes evident that it is both informed and motivated by the interests of the agent. In this case, the fly's disappearance is judged as a necessary condition for the satisfaction of the person's interest in eating. Thus the element of necessity is strictly conditioned by the goal being pursued. In this way the value of the goal determines the value of the means. The motivation to be rid of the fly will depend upon the value assigned to the act of eating the sausage. A hungry person with a taste for barbecued sausages will assign the fly's life a thoroughly negative value. Another, who is not especially hungry or prefers a vegetarian diet, will be much more tolerant of the fly's presence.

If killing flies hardly seems a moral issue, we need only remember that human beings routinely kill other human beings as if they were pests. Eradication is then conveniently rationalised as a simple matter of 'necessity' in protecting the livelihood of the community. In the practice of slavery, different ethnic groups are regarded literally as beasts of burden, so that their treatment befits their lowly status. Here is the constant danger that lies in attributing a factual status to moral claims. If such claims are based on facts, their force cannot be ignored or disputed as judgements of value can. This is precisely the appeal of the categorical imperative in purporting to provide an objectively valid criterion of moral judgment. But it also shows why it is a false and dangerous idealisation of human reason to insist that it can provide an unconditional foundation for moral action.

As they are founded on judgements of value, moral claims are inherently subjective. Nevertheless, they are also objective in that they express values about actual or possible conditions and events. In the process of selecting and devising the best means of achieving our ends, we cannot avoid the necessity of assigning values to objects and events. This relative necessity by which we invest the world with value is subjective to the extent that it is directed by personal beliefs and desires. But this subjective element in any question of judgment must also be balanced by an element of reasoned justification that takes account of the objective properties and events that have informed the judgment.

Necessity does dictate that some facts about human nature and the physical world are inherently value laden insofar as they are biologically beneficial. For most healthy organisms, food and sex are in fact desirable. Such basic desires and needs are biological facts that serve the organism's physical well-being. At the same time, however, genetic variation combined with autonomy, ensures that these instincts can be subjectively valued in the form of distinct preferences e.g. for celibacy or vegetarianism etc. Thus, these objective, life-enhancing values have significant room for subjective evaluation and expression. Food is necessarily and demonstrably good for well-being. Therefore, it is perfectly valid to assert that human beings ought to eat if they are interested in their well-being. This biological imperative is so essential that it scarcely seems hypothetical. Yet, precisely because it is a condition for living that can be revoked, it would not justify the categorical claim that all persons have

a duty to maintain their lives, let alone improve them. In this strict sense, it is indeed fallacious to assert that any given condition or way of life is necessarily good. The fact that food and sex are necessary for survival and reproduction does not mean that these ends are good. Like necessity, goodness is not a property that resides in objects or events. It is a judgment of their value in relation to a goal. Therefore, goodness is a matter of personal preferences and can never be settled by appealing to nature. It is up to each individual to judge the value of different means and ends available to them.[129] But to the extent that these possibilities are only made practicable by certain conditions of life, they do have an objective justification grounded in utility. Although a conditional basis for moral value, utility is the one imperative that must be observed by any animal with an interest in its well-being.

It might be objected that such a conditional basis for moral judgement and action gives an inordinate latitude to the range of behaviours which could theoretically qualify as morally defensible. Thus, it might be thought that a person no longer interested in their own self-preservation would therefore no longer be bound by the hypothetical imperative to respect any other person's desire for self-preservation. However, from the fact that *A* is no longer logically bound to protect his own well-being it does not follow that *B*'s interest in *his* well-being no longer has the *moral* status of a hypothetical imperative. Precisely because a hypothetical imperative to do *X* only binds those for whom *X is* a chosen end, it has no authority over those who have *not* chosen *X* as an end. As such, the self-destructive person *A* is *only* released from a hypothetical commitment to *his own* well-being, *not B*'s. Therefore, there is no logical reason to suspect that imperatives which are conditional upon agents' concern for their well-being might somehow lose their moral status, merely because someone else does not share that concern.

Whether or not any given measure of utility is good must remain an open question. As a philosophical point, Moore's challenge is unassailable. However, the human mind did not evolve by grappling with abstract conceptions of value. Rather, it is the capacity to reason instrumentally that accounts for the development of human intelligence. Interpreting phenomena in terms of their *general* utility is what enables us to function successfully. Therefore, *in practice* the value of utility is not an open question for a mind that cannot avoid having to evaluate objects and events in terms of their pragmatic effects. Indeed, the very concept of 'value', however we attempt to define it, loses all meaning if it does not refer to such pragmatic utility.

Many people would find it insulting or deplorable to suggest that utility is the universal principle that governs their lives. If all judgments of goodness are ultimately judgments of utility, then even the most exalted human values and ideals may seem to be either a sham, or irredeemably sullied by self-interest. While agreeing that utility is a basic principle of thought and action, we are inclined to regard many ethical values and ideals as being explicitly opposed to utility. An entire tradition of thought from Plato to Kant, along with the major religions, has been dedicated to living according to a 'higher' source of meaning that commonly expresses disdain for material goals that are merely useful. However, it has already been observed that the possibility of unmotivated value is denied by the fact that human behaviour is essentially goal-directed. The ideals expressed in art and religion may not be so concerned with material ends. In this respect, they are genuinely alternative ideals, but ideals nonetheless. The idea that art can be pursued 'for art's sake' has romantic appeal. But as objects of art

need minds to give them value, they can only be created for the sake of an audience, even if it is limited to the artist alone. As Dewey notes, the notion of aesthetic value refers "to experience as appreciative, perceiving and enjoying. It denotes the consumer's rather than the producer's standpoint".[130]

As art and religion primarily serve psychological rather than material human needs, we may be deluded into believing that they somehow transcend the utility principle. To describe a product of the imagination as "useful" can seem like a travesty of its value.[131] Similarly, few people would concede that their cherished ideals are ultimately expressions of utility. It is as if utility properly applies only to concrete acts and is therefore too mundane to reign over the free world of the imagination. Heidegger epitomises the spirit of this romantic revolt against valuing all phenomena in terms of means-ends. But his quest to recover an abandoned "ethos" to replace the "technical" mode of thought has only produced incommunicable experiences of "being" that have no prescriptive value.[132]

Heidegger's life and work are a testament to the impossibility of reconstituting the essence of being human. Although some normative beliefs and values can change dramatically in a single generation, they do so within the limits imposed by the evolved constraints of pragmatic reasoning. Revolutionary changes can even occur as long as they work within the conceptual scheme that has already been moulded by natural selection. Therefore, it *is* feasible to rethink the validity of long held concepts and beliefs, or even to attempt a "revaluation of all values", as Nietzsche had planned.[133] But it is *practically* impossible to redefine the very meaning of the concept 'value' without reference to the identification of means and ends as constitutive concepts in the teleological functioning of reflective agents. Similarly, it would be infeasible to dispense with whole idea of value as if it were an error to believe in means and ends. The architecture of the imagination is founded on these primitive action-guiding concepts. As such, they are vital elements of human psychology and behaviour that have become embedded in our genes, given their durable fitness in allowing us to envisage the likely costs and benefits of feasible actions. Therefore, absolutist and deontological accounts of normative ethics, such as those proposed by Plato and Kant, must be rejected as impractical in prescribing standards of judgement which require agents to ignore or discount the likely effects of their actions. If agents are to successfully envisage the extent to which different means and ends may affect the well-being of themselves and others, they must also learn to rely on a range of basic character traits to guide their judgement. Accordingly, the following discussion will show these virtues to be essential motivational constituents in any stable moral system, as they enable agents to develop the autonomy needed to exercise reliable control of their responses to moral situations. Nevertheless, the analysis will also question the extent to which a viable moral theory can evolve and flourish solely on the basis of such essential virtues, given that their cultivation and maintenance also requires a favourable   social environment

(ii)   The Virtues of Character

As products of the imagination that are assigned to objects and events, values themselves should not, of course, be regarded as genetically inherited. Although our personal values may be 'coloured' by those of our parents, often enough we are at odds with them. In laying claim to our own

personal values, we assert our autonomy. It is this capacity for autonomous self-expression that is genetically inherited, along with a range of broad character traits which still require a conducive socio-cultural environment in which to develop effectively.

For a long time the nature/nurture debate was conducted as a war of opposing ideologies, with both sides committed to decrying the other's experimental results. In such a climate, genuinely scientific research suffers. However, recent advances in many different fields are at last leading to a consensus view that human behaviour is produced by both genes and environment working together. Yet this had been Darwin's idea all along. Modern biologists have been so entranced by the wonders of the gene that they have underplayed the evolutionary value of the environment. Genetic determinism would require that in a relatively unchanging environment any shift in behavioural norms must be due to genetic variation of some sort. Whether by random mutation within the population or through contact with outsiders, or simply as a result of "genetic drift", a new combination of genes may produce a change in behaviour that is better adapted to the environment.[134] Over the generations those born with the new trait will proliferate to the point where its expression becomes the norm, supplanting the earlier one. But while the gene may in this case be the observable impetus for change in a stable environment, this does not mean that the features of this environment are any less forceful in the selection of viable traits. For instance, a sandy desert that has barely changed over thousands of years can accommodate only those variations that are highly adapted to such conditions. Therefore, it is as much the *environmental conditions* in which a novel trait appears that will determine its viability.[135]

Considered in this light, the entire spectrum of moral action can best be understood as the co-evolution of a range of character traits that are selected on the basis of their utility in maintaining the well-being of a group. The use of the term 'group' here should not be taken as an endorsement of any form of group selectionism. As argued in the previous chapter, any attempt to account for the extraordinary diversity of phenotypic traits in humans cannot be captured by models which regard such traits as belonging to a group, rather than the individuals whose behaviours actually express those traits. Thus, in human evolution such terms as 'group', 'community', 'population' etc., should only ever be understood as designating a collection of individuals exhibiting a variety of characteristics adapted to a given environment. Depending on the level of economic development, different moral virtues will differ in their objective value for a particular community. This is why primitive subsistence societies today follow the same egalitarian ethic that is thought to have been practised by the earliest human groups.[136] In a small group, the prospects of survival are greater if scarce resources are shared. Thus, competition within the group is suppressed by the custom of food sharing. The success of the custom ensures its enduring selection as a group norm. Through the application of practical reason the group members begin to exercise control over their well-being. By actively determining the viability of their actions, each individual participates in the selection of the norms that will guide them. In this way, cultural selection further enhances autonomy and reduces the influence of natural selection in determining the conditions of a good life.

The egalitarianism expressed in the habit of food sharing is given explanatory support in terms of the natural selection of certain useful traits that are present in many other species.[137] As an extension of the parental instinct in feeding offspring, communal food sharing does not require any deliberative

planning. The leader of a wolf pack need not be aware of the benefits of sharing its kill. The habit is entirely instinctive and mechanical, requiring only the disposition to respond appropriately. Yet, without the disposition to share, no form of egalitarianism can become selected by either nature or culture.

Aristotle recognised that the good life can best be realised through the cultivation of the natural virtues of character. Like other animals, human beings are naturally equipped with the capacities required for achieving the ends that promote their well-being. Both practical intelligence and a range of character traits are differentially expressed in the teleological functioning of animal behaviour. Animals of some species are more aggressive than others. Some are more cautious. Between the paradigms represented by the Hawk and Dove, there is a range of temperaments that will be displayed in different strategies of defensive behaviour. Eventually, selection pressures will favour the evolution of a dominant trait that comes to characterise a species' behaviour. In a mouse, for example, timidity is a natural virtue for its way of life. By contrast, a mongoose's success in killing snakes depends upon a lack of fear that would be dangerous for the mouse.

Among the earliest humans, such traits as courage, prudence and temperance would have been favoured by natural selection, as they are highly useful in the struggle for survival. A courageous disposition is vital for success in hunting and defence, while temperance and prudence promote the economical use of resources. Thus the evolutionary history of our species accounts for the origin of these natural virtues. However, these character traits could not be regarded as moral virtues if they merely dispose an individual to perform an act. A bird defending its chicks cannot be praised for its parental instinct. But in our own species this same natural instinct becomes a moral virtue that is normatively commended as a duty. An instinct that is naturally valuable becomes morally valuable for the same reason. The natural "is" that comes to be accepted as the moral "ought" occurs as a result of cultural selection taking over from natural selection.

Morality can thus be largely understood as the artificial selection of those life-enhancing actions that were previously enforced by nature. Once the human brain had fully developed the capacity for inductive, instrumental reasoning, human evolution was no longer at the mercy of nature. The increasing awareness of our ability to control our own actions meant that they could no longer be determined solely by our instincts. With reason and autonomy reducing the effect of natural selection, those natural virtues that had once been vital would need to be consciously developed and directed toward the common welfare of the group. Previously, a generous disposition would have been enough to elicit the practice of food sharing. However, as self-awareness and instrumental reasoning grows, so too does self-interest. The generous person will be cunningly exploited by those who lack this virtue. As a consequence, the egalitarianism that had evolved through the natural selection of generous behaviour is undermined by the advancement in practical intelligence and autonomy. A natural virtue that had once served the mutual needs of the group could now be manipulated for the satisfaction of others' selfish ends.[138]

Virtuous behaviour that had once been natural would now have to be artificially produced to reduce competition for resources. Thus the natural virtues of prudence, temperance and courage must be actively developed and directed toward those ends that constitute the good life. Aristotle recognised

that practical intelligence is necessary in determining the appropriate means of attaining those ends. This is the level of voluntary control that operates in the goal-directed motion of most animals, even the fly. But this intelligence is limited to the capacity to recognise and respond to the means and ends that are appropriate for the animal's way of life.[139] It is only in humans that this practical intelligence can be applied to the choice of ends as well as the means to realise them. With the aid of practical reason, natural virtue is educated by a conception of the good life and is thereby made moral.[140]

Moral education is the means by which culture replaces nature in the promotion of co-operative social behaviour. In this respect, the human race owes its continued existence to the artificial reproduction of virtue. However, without the influence of moral rules, virtues can become vices that are inimical to well-being. Traits that had been selected for their survival value retain that value so long as survival is paramount. Courage is a virtue in adversity. But the same trait loses its utility for an individual living in a comfortable, protected environment. In a community with abundant resources, temperance is no longer a virtue. Capitalism is positively driven by *in*temperance. Under such conditions, prudence becomes absorbed by a suspicious, self-centred conception of goodness that discourages co-operation.

The natural virtues that would lead primitive human communities to develop egalitarian practices have now become adapted to artificial environments that evolve by cultural rather than natural selection. Laissez-faire capitalism effectively mimics the process of natural selection by promoting the welfare of those who 'win' in the 'competition' for resources, while the rest are punishable for being unfit. Thus the virtues selectively favoured in a capitalist culture are those that a 'selfish' gene would easily parasitise and corrupt: courage turns into aggression, prudence into suspicion, while temperance is a loser's virtue, only fit for the poor.

Of the four cardinal virtues, prudence, temperance and courage are natural dispositions that must be cultivated into serving moral ideals. This is precisely why they are vulnerable to corruption by the cultural environment.[141] Justice, however, is not a virtue that occurs naturally. As Hume observed, justice is a human invention, a virtue that is artificially constructed without the natural origins of the other cardinal virtues.[142] But the beauty of justice is in its construction, which is deliberately designed to resist corruption. The truly just person cannot be manipulated into pursuing selfishly competitive goals. Undoubtedly, this is how justice evolved, as a means of checking competition when it threatens the stability of reciprocal altruism. Effectively, justice functions as an artificial scheme for publicly maintaining the stability of reciprocal altruism. The requirement of balancing competing interests is a powerful device for delivering a conception of the good upholding an egalitarian ethos that would otherwise be only precariously maintained.

Honesty and impartiality are not moral sentiments that are based on any innate dispositions. Courage, prudence, temperance, generosity, self-sacrifice, modesty: such virtues are readily accountable as natural traits that are observable in the absence of any moral rules. Animals within other species display similar variations in temperament with little or no social enforcement. There is no question of anthropomorphism here. Aristotle is right to insist that the 'courage' of a lion should be understood as a metaphorical description.[143] At the same time, however, he admits that the courage of a man and a lion both partake of the same instinctive basis and have the same natural function. The lion's

courageous display is comparable to that of a child who has not yet learned to control its emotions.[144] Determined to keep a toy in its possession, a child will scream with anger. For the same reason, a lion with a fresh carcass will roar at prospective scavengers. In time, the child will normally develop an awareness of itself and others that refines and directs the expression of anger within limits that are socially acceptable. In this way, the natural virtue of anger is transformed into the moral virtue of courage. But both the emotion and its function remain unchanged. What is a virtue for the lion and the child is a virtue for the civilised man, whose courage has the same goal of self-preservation. The man beating his wife in a jealous rage, or the soldier defending his ground both act out of the 'courage' of their convictions. Both *believe* they have morality on their side, giving them the right or duty to protect their own interests.

Examples such as these show the dangers of a morality focused on self-preservation. By definition, a moral virtue can only be used for the good, or it must be deemed a vice. But as some of the traits that are held to be virtues can also be vices in the interests of self-preservation, then these traits are either not genuine virtues, or the Aristotelian conception of moral virtue is an unreliable guide to good action. In claiming that the virtuous person's practical intelligence will inevitably lead to the right moral decisions, Aristotle only creates a viciously circular rationale that identifies virtue only in hindsight. Can practical intelligence be trusted to find virtue between the extremes of vice, where self-preservation is the primary concern? Depending on the situation, a person's propensity for courage can fluctuate between the extremes of rashness and cowardice.[145] The emotions of a man who beats his wife at home may also have worked to save the lives of strangers in wartime. Similarly, what seems a clear act of cowardice from one perspective may be courageous from another.

When the American President, Harry Truman, gave the order to bomb Hiroshima, did he act out of rashness, courage, or cowardice? At the time, it would have been scarcely possible to pass judgment from any perspective, least of all Truman's. But even now, over fifty years later, does the historical legacy of the act permit an objective assessment of Truman's moral character?[146] With the benefit of hindsight, the costs and benefits of the act can be seen more clearly. From a utilitarian perspective, we can attribute a degree of responsibility to Truman's powers of judgment. After weighing up a wealth of historical data on all fronts, we may confidently conclude that the bombing so hastened the end of the war that it prevented more suffering and death than it caused directly. But even if history should prove that Truman made the right decision at the right time, that need not be due to the unfailing virtue of courage. For all we can know, it may have taken yet *more* courage to delay giving the order.

Judging the morality of an act in terms of its consequences would seem to provide an objective measure of responsibility that can be witnessed and endorsed by all. But any attribution of moral responsibility must also consider the extent to which consequences can be foreseen and weighed by the agent. After all, it is the agent's state of mind in evaluating the situation that precipitates the action. Before we can judge whether Truman made the right decision, we need to assess his capacity to do so. An examination of the consequences of that decision can only ever be a reassessment after the event. Furthermore, if Truman's decision had been shown to be rash or otherwise immoral, it is not

only his judgment that would need to be reassessed. All those who *trusted* his judgment at the time would be shown to be poor judges of character.

Notwithstanding the indispensability of moral virtues, the above difficulties substantially undermine the reliability of moral virtue theories in furnishing independent guides to moral judgement and action. In the absence of any public principles, procedures or norms by which such character traits can be regulatively directed, they are liable to be misdirected or corrupted. If virtues are to remain *trustworthy* guides to moral behaviour, the probable costs and benefits of their practice in any co-operative scheme must be *mutually* assessable by all of the individuals affected. To sustain the level of reciprocity essential to the evolutionary stability of political and moral systems, a corresponding level of transparency and accountability is needed to protect individuals from the dangers of exploitation. Therefore, the next section considers the evolutionary role of justice in balancing competing interests which would otherwise diminish the mutual rewards of co-operation.

(iii)   Justice and Reciprocity

Deciding the moral value of an act after it has been performed is like picking the winner of a horse race after the finish. However, observing a horse's performance may be a useful guide to its future performance. Evaluating a person's moral behaviour typically follows the same procedure. Throughout our lives we continually invest our trust in others. Reciprocity cannot even begin unless one party is sufficiently willing to trust that the other will be similarly motivated. But on what basis can strangers judge each other's intentions with neither knowing the other's true character?

Adam is lonely and tired of his bachelor life. He wants to settle down and have children. Eve is newly divorced with two small children to care for. Having both suffered from failed relationships in the past, neither is willing to make the first move in any encounter. At the same time, however, both are also beginning to despair of ever finding a suitable partner. Now, if Adam and Eve were to meet by chance, as perfect strangers, there is little information on which they might reliably assess each other as prospective partners. The dynamics of sexual attraction have evolved to ensure that partnerships will be formed, if only for brief periods. As Adam and Eve are interested in forming a lifelong stable relationship, sexual attraction is a weak foundation for building the trust and fidelity that they both seek. While sexual satisfaction is a vital factor in sustaining such a relationship, it can be sought with a variety of partners, for as long as they are freely available. Therefore, it is even more vital for Adam and Eve to assess each other's character accurately. Even if it is lust that brings them together, it is trust that will keep them so.

Both need to know that the other can be trusted to fulfil their expectations of the relationship. If the partnership is to be mutually rewarding, both parties must have a clear understanding and acceptance of what each hopes to gain from the arrangement. Its success will depend upon the degree to which both parties' expectations are correlated.[147] Only then can the terms of the arrangement be fairly negotiated on utilitarian grounds. For this reason, the imposition of a "veil of ignorance"[148] that disguises the interests of contracting parties is the worst possible approach to making agreements that are just. The less I know about your motives, the less I will be inclined to trust that they are benign, let

alone benevolent. As a means of equalizing the bargaining power of all parties, the veil of ignorance is intended to induce a just agreement by preventing self-interest from securing any unfair advantage over others. But being ignorant of everyone else's interests also eliminates any possibility of making a reliable assessment of those mutual interests that the contract is designed to deliver.

Justice is achieved by the equal consideration of interests. Discounting the importance of interests denies the possibility and purpose of justice. How can people benefit one another if they are ignorant of each other's needs and desires, their virtues and vices? If Adam and Eve were to get married after a single blind date, they could not expect to remain 'blissfully' ignorant. If they happened to find some happiness together, ignorance will have played no part. It will be due to some fortunate correlation in their needs and desires which they soon discovered. But no-one would be surprised if such a marriage turned out to be a bad mismatch that benefited no-one except lawyers or police. Was it wise for Eve to be ignorant of Adam's violent temper? Or Adam to be unaware of Eve's heroin addiction?

Where is the justice in a contract whose terms do not address the virtues or vices of the persons whose long-term interests are at stake? To begin with, the most basic of all the common virtues - prudence - must surely be exercised as a formal prerequisite for any binding contract. An arranged marriage or a shotgun wedding may be consented to, without reservation, by both partners. But if either partner has given consent imprudently, without a genuine understanding of the other's true character, then the terms of the contract cannot be justly applied. A minimal requirement of justice is that due consideration be given to the *interests* of the parties involved. But if the conventional paradigm of justice insists that it should also be *blind* to the weight of those particular interests, it acts in no-one's interests. In seeking a universal, rational standard of fairness, justice has traditionally been based on the principle of impartiality. By deliberating in ignorance of the particular virtues and vices that influence moral decisions, impartiality is well served. But it delivers only rough justice. Whose interests are served by a principle that is blind to the virtues which *promote* moral interests? Ironically, by ignoring the moral capacities of citizens, a system of justice that is strictly impartial is in that respect *biased* against such factors in judging questions of individual desert. As an old proverb warns, "In the country of the blind, the one-eyed man is king".[149]

As the "constant and perpetual will to give everyone his due",[150] the ideal of justice demands a level of honesty and impartiality that is at odds with dispositions that have been inherited for their survival value. Genetic self-interest is poorly served by such traits, especially in conditions where dishonesty is continually rewarded by reproductive success. Therefore, it is hardly surprising that, as Hume noted, such virtues have no basis in human nature.[151] This explains why justice is an ideal that is hard won and routinely flouted, requiring conventions and institutions to enforce it. From a biological perspective, the concept of justice would seem to be an impediment to self-preservation. In the competition for resources, an individual who is resolved to give everyone his due is liable to be duped and drained of resources by those who are not so disposed. Yet, from within this competitive milieu, the concept of justice emerged in human societies everywhere, sustaining the belief in the value of co-operation against competition.

As long as the belief in justice remained inimical to self-preservation, it would not have survived. Instead, in belief, if less so in practice, justice has become what Gibbard terms a "norm of governance",[152] indicating its widespread community acceptance. The will to give everyone his due is *not* applied constantly and perpetually by many people in any society. But it is the collective confidence in the institutional force of justice that gives its appeal to private citizens. As long as enough people have good reason to believe that justice will protect their interests, they will be encouraged to invest those interests in each other.

Before justice evolved to become a norm of social governance, relations of mutual exchange were negotiated according to rules established by custom or convention. Many of these customs worked to reinforce existing inequalities in social status, where a person of low standing endured the poorest standard of living in the service of those with abundant wealth. The terms and conditions of master/slave relations are scarcely negotiable when the survival of one is entirely at the mercy of the other. Thus, in feudal societies there is no concept of a free and fair contract, as the norms of governance are strictly determined by the power and authority given by social inheritance.[153] The most powerful institution, that had proclaimed its service to the needy, was by now complicit in their domination. Instead of providing support for those in need, the Church allied its institutional power with those possessing the greatest wealth and privileges. In this way, the clerics and the nobility had themselves forged a convention for mutual benefit. Both were able to maintain their wealth and privilege by exchanging favours.[154]

In many alliances, the terms of reciprocity evolve by custom and convention, without any predetermined rules of contract. The rules that come to govern the arrangement are produced by the bargaining process itself. Hume provides a suggestive illustration of this process, in his description of two oarsmen who combine their movements to greater effect without any formal promise or agreement being made.[155] Mackie develops Hume's example to present a more detailed scenario that assumes both men to have the same objective of crossing a river, but that each is unwilling to work for the other's benefit.[156] Here is a literal presentation of the 'free rider' problem. Both hope to reach their objective at minimal cost to themselves. Both also know that each would achieve that goal more efficiently by rowing together. But neither can be sure that the other won't slacken his strokes at some point, in an attempt to reduce the cost to himself. How can one rower dissuade the other from reducing his contribution to the task?

Assuming that the desire to cross the river is equally pressing for both men, it is in their mutual interest to row in harmony by synchronising their strokes. But if *A* is less committed to the task than *B*, *A* may attempt to shift more of the burden into *B*'s hands by slowing down, forcing *B* to row harder. The more *B* allows *A* to slacken off, the more *A* will be encouraged to do so, to the point where he may drop the oars entirely. In such circumstances, where goals are best achieved by co-operation, it would seem that those who are less willing to co-operate will ultimately dictate proceedings. As long as they are unlikely to suffer any great cost by withdrawing their contribution, they will have every reason to do so.

In terms of rational self-interest, each person's motivation to contribute will be largely determined by the expected pay-off in each case. If we imagine that Hume's two rowers have an equal

investment in crossing the river as quickly as possible, then each man can be confident that the other will do his best to keep to the rhythm. If *A* begins to lose his interest or concentration, or tends to be lazy, *B* can make an heroic attempt to compensate for *A*'s declining effort. But then, as long as *B* can maintain the pace to *A*'s satisfaction, *A* has no incentive to row any faster. By increasing the rate of his strokes, *B* only reduces the need for *A* to co-operate. As Mackie suggests, the most rational strategy is *not compensation but reciprocation* i.e. *B* should respond to *A*'s slower strokes with even slower strokes.[157] These "reciprocal sanctions" issue a warning that any withdrawal from the agreement will not be rewarded but repaid in kind. So whenever *A* eases the tempo slightly, *B* immediately does the same so that *A* is constantly reminded that reneging on the agreement cannot pay. Thus *A* is persuaded to resume rowing in rhythm with his partner, the convention is maintained, and both achieve their goal with maximum efficiency. Through this process, Mackie's analysis shows the notion of reciprocity to be operative at a more intimate and immediate level of personal interaction than Hume himself had envisaged with his original example.

The strategy of reciprocal sanctioning succeeds by appealing to the rationality of self-interest that governs the choice of action. Thus *A* and *B* can influence each other's chosen behaviour in a rationally dialectical fashion, until both settle into a mutually acceptable convention. However, this bargaining process need not assume that any participant should be fully rational or so self-interested as to exclude the interests of others altogether. In its primitive form, game-theory can provide a reliable model for interpreting and predicting the decisions of agents who are consistently and maximally driven by rational self-interest. But such models are only reliable to the extent that rationality and self-interest are normatively applied by each individual. Any motivational differences in their beliefs, desires and cognitive powers are selected out of the decision-making process in advance. It is assumed that pre-established norms of belief and desire are universally applied with equal force in all situations, guided by an equally rational efficiency in calculating probabilities.

Subsequent advances in computer processing have made it possible to develop more sophisticated models that present a greater diversity in the distribution of individual traits. Instead of having only Hawks and Doves encounter each other in a static environment, evolutionary game theorists attempt to model more complex behaviours that reflect the interaction of a variety of traits, expressed in different degrees and under different environmental conditions.[158] Hence these models provide a much more realistic representation of the dynamic influences that lead to the evolution of co-operation as a stable strategy. Political scientist Robert Axelrod devised computer programs to test the viability of competing strategies in repeated encounters between two players.[159] At each encounter the only options are to co-operate or defect. Yet from this simple dialectical procedure, a complex series of manoeuvres can be constructed to form a coherent strategy. In this way, a range of strategic rules are tried and tested against each other. The results of these round-robin tournaments demonstrated conclusively the superiority of a principle that is tantamount to the propagation of life - *replication.* The strategy of "Tit-for-Tat" proceeds by simply copying the other's action at every turn after initially co-operating. Thus, even the first move in favour of co-operation is an attempt at self-replication through the agency of another compatible replicator. If the latter fails to reciprocate at any stage, that

defection signals its reproductive incompatibility to the Tit-for-Tat player, who then responds by also defecting.

It should be noted that the term 'strategy' here is not intended to imply any especially higher order of intelligence or reflective planning in such models. Like 'altruism', the term is routinely used in evolutionary game theory to describe a form of behaviour which may be guided by different degrees of intelligence. Considering the number and variety of complex algorithms that were tested against Tit-for-Tat, it could be conceived as a successful paradigm for social behaviour in all species possessing a modicum of intelligence. It employs the simplest two-step decision rule to yield the most reliable long-term benefits for survival. Yet, if it is the best recipe for survival, why does it occur so rarely in nature? In the natural world, self-preservation is seldom achieved by the reproduction of a fixed response. An organism whose entire social behaviour is governed by a single universal rule will benefit only insofar as its neighbours also follow that rule exclusively. However, genetic and environmental changes will always work against the fixation of any behavioural trait that is insensitive to such changes.[160] As shown earlier, an animal that has learnt to employ a range of alternative strategies will outlive its rivals when conditions alter.

The Tit-for-Tat rule works by consistently rewarding co-operation and punishing defection. Thus, like justice, it is an artificial selection mechanism that operates internally i.e. through the decision-making process. As a solution to problems of conflicting interests, it has much in common with Kant's Categorical Imperative.[161] To best succeed, its application must be *universal* and *unconditional* in the sense that it cannot accommodate exceptions if it is to effectively maintain its dominance over other strategies. Although its responses are indeed conditionally determined by the behaviours it encounters, the Tit-for-Tat rule is itself *unconditionally* applied i.e. unfailingly applied on each and every occasion where those behaviours are encountered. Insensitive to the variables of character and situation that condition the pragmatic value of any strategic choice, Tit-for-Tat thereby contains the pursuit of private interests at the expense of autonomy.

Ultimately, the principle of Tit-for-Tat is unconcerned with the interests of others. Even when co-operating, it acts on its own behalf by demanding others to follow suit. While relentlessly sanctioning the behaviour of its neighbours, it refuses to be sanctioned by them. Even in encounters with fellow Tit-for-Tat strategists where sanctions need never actually be applied, their mutual responses remain unconditionally enforced in being always bound to co-operate without exception. Thus, the unconditional and universal reciprocity demanded by Tit-for-Tat is incompatible with the practice of reciprocal altruism, where individuals willingly act on each other's behalf, accepting the costs as well as the benefits of co-ordinating their efforts to mutual advantage. The former achieves co-operation by coercion, ignoring the variations in beliefs, desires and capacities that condition the scope of what each person can expect to contribute or receive from a given arrangement.

Tit-for-Tat is a totalitarian solution to the need for co-ordinating the productive activities of citizens who would otherwise fail to do so. Although Tit-for-Tat is itself only a rule which may be selectively employed in human interactions for different purposes, when considered as a stable political principle for engendering co-operation in human societies, its unconditionality and disregard for the welfare of those to whom it is applied can only promote the most primitive conception of justice - that

of *vengeance* and deliberate *blindness* reflected in the irony of the motto "an eye for an eye". If justice is to be a norm of governance that is willingly endorsed by the citizens whose welfare it protects, it must be vigilant in its concern for each individual. As such, it cannot afford to ignore the uniqueness of a person's character and situation, insofar as differences in these factors condition each one's level of commitment to the ideal of justice. Applying the pure Tit-for-Tat principle to the regulation of human societies can only be counterproductive in the end, as obedience is achieved by instilling terror and suspicion into human relations on every level. So it is that totalitarian regimes that espouse communist ideals inevitably dispense with the notion of justice as fairness, to discourage the pursuit of private interests that are seen to threaten the spirit of co-operation. Yet, by replacing the principle of fairness with one of retaliation, the individual's motivation to co-operate is destroyed and the ideal of distributive justice abandoned. "From each according to his abilities, to each according to his needs"[162] is a maxim that recognizes fairness as constitutive of distributive justice. To treat everyone equally is not to treat them fairly. As long as abilities and needs are *unequally* distributed among individuals, treating them all equally denies justice for all.

Like blind justice, Tit-for-Tat treats all of its neighbours equally. Hence, it appears to be strictly impartial in all encounters. But this semblance of impartiality has nothing to do with justice. It is only impartial in that it treats everyone in equal measure, paying each in kind. But for this very reason, the rewards or sanctions it bestows are as partial as those with whom it interacts. If treated ungenerously, it returns the injustice. Thus, it behaves as selfishly as those whose selfish behaviour it is designed to correct. While it succeeds in eliciting co-operation from others, it is unconcerned with the character or motives of those with whom it associates. This inability to interpret the motivational preferences of others makes Tit-for-Tat much more vulnerable to manipulation by Machiavellian strategists misrepresenting their true dispositions. As a result, it is just as liable to heap rewards on those whose activities undermine distributive justice, as it is to ignore those who are wholeheartedly committed to it. A person with anti-social criminal intentions will be more than willing to return the favours of someone who remains ignorant of their activities or long-term goals. By the same token, those who are hesitant in responding to the demand for reciprocity may very well be misjudged on that account. For their hesitancy need not be read as signifying any lack of interest, let alone any malicious intent. Instead, such wariness may be a sign of heightened sensitivity and concern for others that has been chastened by some recent abusive encounters with extremely intolerant characters like Tit-for-Tat!

Characters who are potentially well-matched cannot afford to misrepresent themselves for too long. If Tit-for-Tat is employed as a means of testing a potential partner's level of commitment at little cost, it is unlikely to be very successful. It is bound to detect a high percentage of false positives as well as being seduced by the simplest of Machiavellian deceptions. Unwittingly, the most fruitful opportunities may be missed, while corrupt liaisons are cemented instead. Nevertheless, these defects can be remedied without discarding the principal value of Tit-for-Tat. Primed to recognise and return whatever comes its way, it is a strategy that is highly stable and resolute. In that respect, it is precisely what is required of justice as a norm of governance that can be relied upon to treat all comers equally on all occasions. Tit-for-Tat's evolutionary success shows reciprocity to be the most stable foundation

for ensuring co-operation on equal terms. But again, treating everyone *equally* does not guarantee that each is treated *fairly*. To that end, strict reciprocity, as the first condition of justice, needs to be tempered by an overarching sense of fairness while still maintaining its authoritative power. In Kantian terms, the question is whether justice as fairness can be widely endorsed as a moral imperative that nevertheless remains hypothetical. Insofar as fairness is *conditioned* by giving weight to the individual's character, situation and preferences, is its *imperative* value thereby diminished?

As an inflexible response to each and every character and situation, the Tit-for-Tat strategist obeys a categorical imperative that sees only black and white. It prejudges the prospects for a productive alliance in absolute terms by admitting only two possibilities for interaction at each encounter. In decision theory, these expectations are expressed in terms of subjective values that represent the relative importance of preferences. Thus, if Tit-for-Tat is applied to human decision-making, instead of being a blind, mechanistic agent of reproduction, it has the potential to generate commitments to conventions that evolve autonomously, reflecting the agent's own interests. In Axelrod's 'tournaments', Tit-for-Tat did not actually operate as a strategy, as if it were the result of a deliberative process. Programmed in advance, it acts without forethought. Like a squirrel burying nuts, or a beaver building a dam, its actions may be goal-directed without being performed out of any strategic 'intention' to survive. But as the conditional basis for an economical human decision procedure, Tit-for-Tat is a powerful means of restraining the pursuit of self-interested goals at the expense of the common good. As long as the sanctions that it imposes are fully recognised and endorsed by all who are party to a convention, it will avoid any intentional injustices.

As evolutionary game theorists continue to develop more complex models, capable of representing and testing an increasing variety of mixed strategies within different socio-economic parameters, the evolutionary stability of alternative principles of justice can be more effectively evaluated. The preceding analysis of Tit-for-Tat confirms the intuitive assumption that the adaptive value afforded by a stable social contract can scarcely be delivered by a single rigid algorithm. As shown in the preceding chapters, the autonomy of human agency is dependent upon a high degree of reflective plasticity which is fundamentally antithetical to the unconditional adoption of rigid rules and norms. Thus, the terms of reciprocity governed by a viable social contract must be sufficiently *flexible* to accommodate the evolving needs, interests and capabilities of the individuals it protects. This need to protect the autonomy of individual agents, while also maintaining a fair system of social benefits and responsibilities, raises the question of the extent to which individual behaviours and capacities can be justly accommodated within such a system. The final section of this chapter will address this issue by focusing upon examples where evolving relations among individuals can be seen to create mutual responsibilities which can be reasonably circumscribed by the concept of a social contract.

(iv) The Reasonability of a Social Contract

In contrast to blind justice, where reciprocity only serves to restore and maintain the initial inequalities between individuals, a fair system of justice will apply the principle of reciprocity to rectifying those very inequalities. It is precisely in order to achieve true fairness that justice must be

partial to some degree, so as to compensate for the unequal distribution of those assets that enable an individual to participate in society in the first place. With his "difference principle", Rawls recognised that fairness demands equal opportunity from the beginning.[163] Inherited advantages in wealth or social status cannot justifiably be used to advance one's own interests at the expense of those who have been less fortunate. For the same reason, differences in natural talents or capacities cannot be a fair basis for social exchange. As long as such differences allow one party to exercise a significantly greater power over another, they will promote injustices. But this certainly does *not* mean that it is any more justifiable to eliminate such differences. On the contrary, fairness can only be achieved if such differences are taken into consideration, as it is their utility that constitutes the social exchange 'currency'. Indeed, if it were not for differences among individuals, conventions and contracts would not be necessary. If each person were equally capable of achieving their own ends, reciprocity would not have evolved as a normative principle of human interaction.

Unlike a society of insects, where reciprocity is genetically programmed into behaviour, human beings have had to devise our own self-imposed rules to enforce it. Hume's oarsmen achieve their aim by *resolving their differences* i.e. by establishing a minimal level of effort that they can both be equally satisfied with. In everyday life, it is a practical necessity to voluntarily maintain such norms, even if it is done at the cost of restricting one's own liberty so that the liberty of others is equally maintained. Thus self-restraint and a limited concern for others does not require any strongly altruistic intentions, but need only be motivated by an abiding concern for one's own self-preservation. In abiding by a polite convention such as queuing, each individual accepts the need for a self-imposed limitation to the immediate gratification of their personal desires. Our general acquiescence to such rules is, however, always conditional upon their actual and continual observance by others.

Let us imagine Adam and Eve, total strangers, queuing for service in a bank. Both approach the entrance at the same time, each momentarily hesitates, before Adam suddenly indicates his polite intentions by opening the door and allowing Eve to enter first. With this simple gesture, Adam signals his acceptance of a somewhat outmoded norm of politeness that applies especially to men's treatment of women. While such chivalrous behaviour is now often perceived as paternalistic or a calculated manoeuvre in seduction, it may just as well be what it appears to be - an expression of genuine good will and respect. Unfortunately for Adam, although he does not perform the gesture in the hope of gaining any reward for himself, but only wishes to show his willingness to be of some service, his action is likely to be misinterpreted if it is directed towards a woman who does not herself accept the meaning or value of such a norm. Indeed, the fact that such customs have declined to the point where they have lost their normativity in the manners practised by the present generation, only underscores the fragility of values and beliefs that depend upon common if not universal assent.

Undoubtedly, from our evolved conception of justice based on reciprocity, which is both normative and ideal, there is good reason to be suspicious and critical of any custom which purports to eschew any demand for reciprocity. By definition, a social custom is developed and maintained as a result of interactions among a large majority of individuals, which in turn depend upon and reinforce their shared values and goals. As we have seen, values and goals that have no long-term benefits for those who pursue them cannot survive long enough to gain currency that extends beyond cultures and

generations. Unless their practice is artificially enforced by law or religion, chivalrous acts can therefore only persist as long as there are enough women who approve of them. Eve may rebuff Adam's gesture and insist upon opening the door herself, thereby refusing to endorse a practice that accords women a special status that is at odds with the sense of justice which demands equality between the sexes. In that case, Eve and every other woman who regards such acts as subtly reinforcing an unjust distinction in their treatment by men, will actively discourage their performance with a good conscience. To that end, Eve aims to demonstrate to Adam that she is equally capable of opening a door for herself and henceforth requires no special treatment. Alternatively, if Eve suspects that Adam's gesture is an overture in a strategy of seduction, she will likely signal her disapproval by the same show of independence. In the first instance, she rejects a particular normative meaning that she attributes to chivalrous conventions, while in the second, it is the game of seduction or the seducer himself that causes dissent. In both cases, the response is the same but the impression in Adam's mind will be crucially different, depending on whether he is naïvely chivalrous or determinedly Machiavellian. As the former, like Don Quixote he will soldier on, vigilant and eager to please all women regardless of how he is treated in turn. As he does not seek reciprocity with them, indifference or ingratitude cannot deter him from his quest for justice. In the latter case, like Don Juan, he will be equally undeterred from his equally romantic quest to transcend the bounds of reciprocity. In either case, Eve's appeal to justice – as the recognition of her equality and independence - fails because these elements are by no means sufficient for the realisation of justice in a world where the terms of reciprocity are so often unequal from the start.

By the same token, where is the justice in attempting to rehabilitate a Don Juan or a Don Quixote, so that there are no longer any characters whose desires or ideals are stronger than the norm by which all persons' interests are to be judged as equal? After all, it is the unequal distribution of talents and interests in a community that gives reciprocity its purpose. Our normative conception of justice i.e. balancing competing interests, giving everyone their due, would make no sense if there were not these differences between individuals. The unrestrained passions of eccentric individuals may test our sense of fairness and propriety, but as long as their passions are not opposed to the well-being of those with whom they interact, they are no threat to the normative value of reciprocity. Without such characters, a society is deprived of villains *and* heroes. If we purge society of all *potential* villains, we preempt the appearance of its heroes. Do we expel or ostracise the troublesome mavericks so that only compliant conformists will proliferate? In that case, before too long we will be left with a docile population that is stagnant and defenceless against the most ruthless and sadistic opportunists.

History tells us that for every Hitler or Stalin we need a Churchill or Roosevelt. Whether in defence against a common threat, or simply in the service of our own private ends, however modest, we need a society that encompasses a broad range of character traits that are adapted to the optimal performance of many different tasks. This general utilitarian fitness applies across the entire spectrum of human relations as a determining factor in the selection of desirable traits. Thus, after inheriting the presidency Truman was eventually elected on the basis of his perceived capacity to advance and protect the interests of his nation's citizens. The extreme responsibility of such a powerful position entails the selection of a set of highly developed virtues best fitted for leadership. As Aristotle stressed,

temperamental traits like courage and prudence must be guided by a perspicacious intellect that is able to interpret the unique complexities involved in the situation at hand. In the same way, the desirable traits of friends or partners are selected on the basis of their favourability for providing a high level of mutual trust and concern for each other's particular interests and needs. In transcending such normative boundaries, the Dons - Quixote, Juan (and Bradman, of course) remain rare, legendary figures. Fact or fiction, in excelling beyond our normative expectations of reciprocity, these 'tall poppies' make claims that the rest of us are either unwilling or incapable of accepting. In this sense, it is a truism that the principle of reciprocity ensures that, more often than not, we get the leaders, lovers and friends we truly deserve. For, by repeatedly favouring the degree to which character traits are expressed, we each contribute to determining their value in social exchange.

Consider the example set by the legend of Don Juan. As the archetypal 'seducer', his comportment towards women is represented as selfish and manipulative in the extreme. Driven by his insatiable appetite for sexual gratification, he ignores the conventional terms of reciprocity. As one who devotes his life to preying on others, he is portrayed as a pure Hawk who is unburdened by any moral conscience. As such, his relations with women are shown to be casual and careless. Desiring only to consume them, he cares nothing for their happiness. Yet, as Kierkegaard's analysis reveals,[164] this demonisation of Don Juan misrepresents the ethical problem posed by his mode of life. Far from being unconcerned about her happiness, he is passionately absorbed by the object of his affections. To be sure, he *is* totally unconcerned about the convention of reciprocity. If he scorns reciprocity, it is not out of a selfish refusal to return his lover's affection. Rather, it consists in the refusal to allow that passion to be dulled and diminished by subjecting it to reflective assessment. Kierkegaard rightly distinguishes chivalrous love that is essentially faithful, from sensuous love that is faithless.[165]

As Don Quixote's faithful love is subject to doubt and seeks to be requited, it is under constant reflection, thus "psychical".[166] By contrast, Don Juan's love is free and effortless, governed by a pure desire seeking satisfaction. As such, he has no interest in his love being requited. In this sense, Don Juan is a perfect example of the "wanton" character described by Frankfurt.[167] He acts in order to satisfy his desire without making any evaluation of that desire. For this reason, he is not in fact the archetypical seducer who plots his way to a woman's heart with flattery and deceit, while only seeking sexual gratification. Don Juan does not have recourse to strategies and intrigues designed to misrepresent his true intentions. On the contrary, he fully expresses the strength of his desire and it is this very strength that gives him such power of attraction. It is the power of his *genuine* love for them that arouses a like response in those he desires. If he can be said to seduce them, then he does so honestly and directly, without forethought. As one who loves wantonly, he is not subject to moral condemnation. As Kierkegaard notes, this "is not because Don Juan is too good, but because he simply does not fall under ethical categories."[168]

Insofar as he gives himself completely to every woman in possessing them, Don Juan necessarily finds himself engaged in a reciprocal relationship, one as primitive and blind as Tit-for-Tat.[169] The prodigious number of his 'conquests' is routinely cited as symptomatic of a pathological quest for ever new experiences. Likewise, his lack of commitment to any one woman is interpreted as a fear, as if he were always anticipating some threat or disappointment to ensue in every affair. But if that

were the case, he would be driven to despair upon every new encounter to the point where he would renounce women altogether. Instead, he embarks upon every new affair with the same boundless passion as before. He is one of Camus' "absurd" heroes who lives in the moment, unmoved by past disappointments or doubtful hopes.[170] Without hopes or regrets, he desires and desires only to desire, without ever doubting its value. But although desiring to have a desire does constitute an evaluation,[171] it has no ethical meaning for Don Juan's consciousness. As long as that consciousness remains fixed upon the beauty of the moment, it avoids any moral judgement. It is not ethics but *aesthetics* that guides Don Juan.

Don Juan's wantonness consists in his willing submission to the power of beauty. He is an unwitting seducer who is himself seduced by the singular beauty that he finds in each of his lovers. But his interest is not aroused by those features that distinguish one woman from another, as would be the case if he were bored with the elements of symmetry that regularly bring pleasure to the senses. It is these common attributes that capture Don Juan's attention and so it is that he craves their replication in each and every one of his lovers. In this manner, he succeeds in having his desire continually sustained and satisfied in experiencing the endless repetition of that universal beauty.[172] Don Juan inhabits a timeless universe of his own creation where normative ethical categories do not apply. Reciprocity and its moral connotations of indebtedness, duty and the like also ensure the preservation of desirable characteristics. But Don Juan has no need of such artificial force to keep him attached to one woman after that love has been consummated. Can passion and love be promised or traded like a durable artefact? How can a person be put under an obligation to *desire* another in future?

In raising such questions, Don Juan's character is placed outside the normative scope of a social contract. As long as he fails to comprehend any ethical significance in acceding to sensual pleasures, he can scarcely be expected to meet the demands of a convention designed to restrain them. Whatever the broader social consequences of his behaviour, he cannot be held responsible for the welfare of all those whose lives he has touched in ignorance of any harm done to them. Given that his affairs are governed by mutual consent, devoid of any coercion, can there be any just cause for moral complaint? Can he at least be held morally responsible for succumbing to the charms of married women? But in that case it is only the women who have broken a contract, whose moral value Don Juan cannot appreciate. Furthermore, in their equal readiness to succumb to *their* desire, those women provide all the more support for Don Juan's willing acceptance of *his* desire.

When he is finally brought to account, the story of Don Juan is seen to be a moral lesson in social responsibility. Eccentricities of character are tolerable so long as they do not undermine the acceptability of normative rules for social intercourse. It is feared that if enough transgressions of these norms are permitted they will lose their cohesive force. If Don Juan were to continually succeed in flouting the conventions that most of his neighbours acknowledge and value, he is liable to seduce others to do so. By becoming acceptable, his character would thereby develop a normative influence of its own, in competition with the more established beliefs and practices. Desired and idealised by the women, his behaviour is literally selected by them, through their engagement with him. This new standard of desirability threatens the acceptability of the old, by which they formed their relationships. It is hardly surprising that the women should then feel cheated or disappointed by their prevailing

relationships, contracted in ignorance of any alternative. By threatening the viability of those established relationships, Don Juan's presence challenges the rest of the male population to adapt their standard of behaviour to meet his. Accordingly, some will endeavour to emulate him, so as to replicate his degree of desirability. Indeed, this emulation of such legendary heroes is a common response in such circumstances. But it is an adaptive strategy that is unlikely to be achieved by enough of the population to engender normativity. Thus, their best strategy is to use their collective force to remove the threat, effectively selecting the maverick traits out of the population by banishing the offender. [173]

By the same process of selecting our partners and friends, each one of us actively contributes to determining the normative value of characteristic behaviours. This enables us to co-ordinate our actions to meet common expectations that are mutually beneficial. Without this common currency of beliefs and values, we would have no means of prejudging the likely behaviour of others in our own community. So much of our social well-being depends upon the accurate co-ordination of individual actions, that few of us can afford to ignore the customs and conventions observed by our neighbours. Given that such norms primarily serve to meet the expectations of a majority of others, they will often enough conflict with our own expectations of ourselves. Ordinarily, where these expectations can be met without much inconvenience, most of us willingly oblige. As long as the benefits of compliance clearly outweigh the costs and do not severely hinder the prospects of having our own expectations rewarded, we recognise that it is in our self-interest to co-ordinate our activities. So it is that we oblige others in the expectation that they will oblige us in the same manner, by following the same rules. Thus a normative code of practice evolves on the basis of mutual trust. Occasionally, this trust will turn out to be misplaced, when some person or group is either unwilling or unable to accept such obligations. But if, as in Don Juan's case, there is no intention to abuse that trust by making any false promises, [174] then there can be no justification for demanding any social commitment from those who do not themselves expect any such commitment from others. If they cannot freely consent to the standard conditions of the prevailing social contract, then they cannot be compelled to, without rendering the contract void. As the practice of individual autonomy is itself one of the goods to be protected by the contract, the freedom to consent is itself a necessary condition of entering into it.

While no single individual can be legitimately conscripted into fulfilling social obligations that they themselves do not accept as legitimate, then, of course, such persons cannot expect to receive the benefits of co-operation. This purely logical point is often cited as one of the major theoretical problems of contractarianism, in that it sets quite strict qualifications to granting an individual's right to social goods. Instead of possessing a natural right to common goods, the individual must make some regular commitment to earning that access. But the terms and conditions are too often non-negotiable, having been already established by those who have assumed the authority to contract such arrangements. However, given that there can be no *natural* right to assume such authority to determine contractual obligations in the first place, then by what right can anyone deny common goods to another? In other words, if rights to goods must be contracted, then how can those in an 'original position' negotiate the *right* to appropriate and distribute goods, yet deny this exceptional privilege to others? This raises a host of further questions concerning the legitimacy of the very concept of rights generally, and their contractual viability which will be addressed in later chapters. But here, in Don

Juan and all the other outsiders he represents, it is specifically the condition of consent which raises difficulties. For there are many who lack the will or capacity to consent to agreements which they either cannot comprehend or simply do not value enough.

On what basis must a social contract preserve the well-being of those who have not acknowledged or endorsed its governance? Most importantly, neither a tacit nor explicit act of consent alone would suffice to warrant such rights. For, even if consent might serve as a demonstration of an autonomous decision, it does not provide any measure of its rationality. Those who cannot fully grasp the commitments and responsibilities entailed can hardly be expected to be responsive to them. It would be extremely unreasonable to demand reciprocity from persons whose behaviour shows little recognition of the concept. But even a full understanding of the meaning of reciprocity need not reflect the degree of moral sensibility that is required to appreciate its social value. Children quickly develop both an understanding and appreciation of reciprocity from their first experience of the world. As soon as the infant begins to distinguish itself from surrounding objects, it gradually learns how to manipulate those objects to change their actions. But in the process, while discovering its own power over objects, it also learns that some of those very objects also possess that same power over the child itself. Thus, the instinctive act of crying evolved to stimulate a feeding response to which the child in turn responds. In short, the child learns how it can act upon others to cause them to act in return.[175]

In the relationship between mother and child, the reciprocity begins at an instinctive and unreflective level. As the following chapter will reveal, both have normative demands and expectations of each other, but for as long as the child lacks a certain level of understanding and self-control, it is only the mother, or other parental figure, who is obliged to meet the child's demands, however unreasonable. It is only in becoming reasonable that they become negotiable, as then they become subject to norms of *practical* reason, which test their social viability. In the meantime, however, the physical and intellectual incapacities that temporarily exclude young children from consenting to parental or societal demands, do not excuse the latter from their responsibility for the welfare of the children. While it must be admitted that there are many forms of relationships, reciprocal and otherwise, which cannot and ought not be subject to contractual arrangements, this does not undermine the value of the contract in its broad application to laws and political institutions. As that value consists in its functioning as a political process for identifying and correcting social injustices, then it need only apply to governing those relations where such injustice exists.

Nevertheless, children and all those whose lack the physical or mental capacity to acknowledge and accept the rights and obligations of a social contract, need not and indeed, *must not* be deprived of the benefits that others can provide for them. For most children, their subsequent development of those required powers of autonomy and judgement will qualify them for consent. In liberal democracies, the notion of a legal 'age of consent' reflects this normative belief in protecting and preserving the child's potential participation in communal life. But most importantly, as future members of society, their well-being is intrinsic to the sustainability of the society itself, which in turn, exists primarily for the self-preservation of its individual members. As this self-preservation and well-being is the rationale for the contract itself, those who are already under its protection are thereby

bound to ensure the same protection of every other member of the society, including those who are unable to reciprocate.

It might be argued that a capable citizen cannot be obliged to contribute to the welfare of others who are not similarly obliged to them. But as Tit-for-Tat showed, if strict reciprocity remains the only agreeable norm for producing mutually beneficial goods, then those goods will soon be in short supply, as so many possible partnerships are rendered non-negotiable by unacceptable differences. A homogenous society of thoroughly like-minded individuals has characterised many utopian illusions that have only betrayed the autonomy of their citizens for their own 'good'. Whatever goods a society freely trades, within and without, must by definition, be selected on the basis of that freedom. To be truly self-governing, all citizens must retain the autonomy to determine what is good for their individual well-being. By this very process, each person inevitably contributes to the well-being of the society through the expression and fulfilment of their own particular needs and desires. In this sense, even those whose physical or mental disabilities greatly diminish the scope of their activities, all contribute to selecting the depth and range of values that constitute social goods. Such diversity of needs, desires and values that each person brings to others should be all that is needed to qualify for the full rights of citizenship anywhere. But this does not amount to an endorsement of rights being natural. While the biological necessity of self-preservation can become the basis for justifying a universal right to material goods, that certainly does not mean that such a right has somehow been authorised by that natural necessity alone i.e. in the absence of any rational and public process for determining the conditions in which self-preservation is necessarily valuable. In particular, the evaluation and distribution of those goods, material and otherwise, can only be the collective result of our own personal choices which a social contract can best protect.

As autonomy is a necessary condition for the realization of our chosen ends, it must also be a precondition for contracting co-operative goals. Therefore, its development is of primary importance to every individual's own self-development, which, in turn, increases the opportunities for social exchange in the community and beyond. As such, the exercise of personal autonomy has a undeniable value for every human being. Even those for whom life has no value cannot fail to value the autonomy that enables them to choose death. As even the desire to limit one's autonomy can only be met through its agency, it has a demonstrable, universal value that is endorsed by our choices and actions. For this reason, it is a self-perpetuating, self-validating principle, deriving its value from the choices and actions it enables. Thus, it is also a prerequisite for the possibility of reaching agreement on those choices and actions that would best allow our autonomy to flourish so as to make further choices and actions. By allowing us to join forces through co-operative choices, autonomy enhances the opportunities for self-preservation and the pursuit of greater goals. The scope for self-development is therefore dependent upon the scope for productive interaction. The greater the variety of character traits expressed within a population, the greater the opportunities for productive partnerships to form. Under such conditions, even the greatest handicap to the expression of autonomy ought not exclude a person from any such opportunities. For, like everyone else, *by virtue of* their limited autonomy, they *need* the opportunities which the autonomy of others can provide.

In this way, one person's express need functions as the environmental stimulus for the expression of those traits in others that are best fitted to respond to that need. The greater the strength and frequency of needs expressed within a population, the greater the adaptive value of developing the corresponding virtues that generate a disposition to meet those needs as directly as possible. Therefore, far from being a liability for their own, or the community's well-being, those lacking the capacity to lead more independent lives directly contribute to the moral virtues and practices that evolved to benefit themselves and those who assist them. Thus, empathy has evolved to evoke a generalised responsiveness to others beyond one's kin, thereby furnishing the basis for the widespread selection of moral sentiments which encourage the performance of altruistic acts.[176]

The strength of support and concern given to those with physical or intellectual disabilities is therefore a measure of a culture's evolutionary success. A society whose members do not tolerate such incapacities only demonstrate *their* own incapacity for the empathy that would otherwise ensure their survival and well-being. A widespread deficiency in empathy is a recipe for extinction, as evidenced by its vital expression in the moral language of all surviving cultures. Though not abandoning all of their moral customs, the collapse of the Ik's moral system, as described by Colin Turnbull,[177] serves as a stark reminder of how precarious is the capacity for empathy when sudden and brutal changes in living conditions leaves individuals unable to provide for their closest kin.

As individuals, we have evolved a characteristic ability to respond to the needs of others, as they have to us. However, as this expectation of mutual aid becomes the norm, the moral concept of "obligation" serves as a rational means of warranting the legitimacy of that expectation. Whenever *A* benefits *B* in some way, *B* is held to be in debt to *A*, regardless of *B*'s evaluation of the situation. The service that *B* receives may be totally unsolicited and little valued. Yet *B*'s lack of appreciation and gratitude is regarded as reprehensible, as is his refusal to accept that he is bound to do *A* some service in return. Furthermore, *B*'s capacity or will to benefit *A* is not seen to diminish or cancel *B*'s debt, unless *A* allows it. In short, by the normative understanding of obligation, *A*'s act has somehow put *B* in his debt, *whether B likes it or not.*

The concept of indebtedness is built into moral thinking because of its importance in ensuring that goods are shared so that as many individuals as possible should contribute to the general welfare of the community. As such, it follows from recognising the principle of fairness itself to have an *a priori* warrant in legitimising the terms of contract. Thus, in Rawls proposal, justice as fairness is a conceptual and procedural norm that is necessarily presupposed in justifying the contractual approach.[178] Equality and impartiality are already assumed as essential features of the initial bargaining position, as specified in the assignment of the ideal agents' two "moral powers" i.e. a sense of justice and a conception of the good.[179] For this reason, it is hardly surprising to find that the two principles of justice ultimately chosen by those agents should be a close reflection of the moral powers that enable each of them to benefit from such an agreement. Among the alternative conceptions of justice, it is those principles affirming the value of liberty and equality that would have direct appeal to agents for whom liberty and equality are at stake, rather than some particular conception of the good of which they are ignorant under Rawls' scheme. In other words, by depriving them of any personal stake that might bias their choice of just principles, Rawls inevitably allows his agents little choice in determining

the scope of justice. This might seem to suggest that Rawls has merely designed the specifications of the original position so that the ideal rational choice is guaranteed to favour the liberal democratic conception of justice.

Given that their task is to determine principles of justice which are universally acceptable, Rawls must allow each individual agent to proceed from a sense of justice and a conception of the good. Quite rightly, as representatives of a 'well-ordered' society, they are already endowed with their own moral beliefs and values. Precisely because such a society reflects or tolerates a wide diversity of individual value judgements, it requires a publicly acceptable standard of fairness to govern social interactions. In that case, the value of fairness and impartiality is indeed built into the moral thinking of agents living in such a society. However, the fact that a prior commitment to fairness and impartiality inevitably leads to the rational choice of liberty and equality as the central concerns of justice, only proves Rawls' point. Whatever our personal interests, practical considerations alone would show those two principles to provide the best means of protecting those interests. This practical solution to problems of moral justification is one of the great virtues of the contract method. By proposing a political justification as the source of moral value and commitment, there is no need to resort to dubious metaphysical claims.[180]

But for this very reason, Rawls should not be content with the *conventional* description of duties and rights as being 'natural', in contrast to the contractual structure of obligations.[181] For, even if these duties would be rationally chosen in the original position and shown to be in "reflective equilibrium" with our *prima facie* intuitions of moral right, that would only confirm their value as social conventions. As such, an individual's duty to aid another is subject to the same conditions of knowledge and consent that make obligations binding. From a contractarian perspective, there can be no justification for universal, involuntary moral requirements foisted upon individuals. Rawls wants to show that while moral obligations can only be acquired through voluntary acts committed by particular individuals, there are certain general duties owed to all human beings which follow from their having the requisite moral powers to recognise and endorse the principles of justice that would be chosen by their representatives in the original position. However, being brought up in a well-ordered society governed by such principles neither guarantees nor legitimates their authority over those who would not choose such principles.

Conceived as the result of agreements between individuals, moral obligations are governed by conventions like promise-keeping. But it is not inherited social practices *per se* that generate the belief that a debt is owed to one or other persons. A well-ordered society is aptly described by Rawls as a "social union of social unions",[182] where interactions are governed by ends that delineate the obligations of different social unions. But these obligations are not set by any act that has already been performed. On the contrary, they are based upon the expectation that an act will be performed by another party, even before any formal norms have been established. Scanlon gives several examples to show that obligations arise primarily from our expectations of "what we owe to each other", which do not depend upon any prior act or agreement. Even an encounter between two strangers in a "state of nature" can cause one to have a justifiable expectation that the other is obliged to meet.[183] *A* has made a poor attempt to kill a deer and his spear has landed on the opposite bank of a fast-flowing river.

Suddenly a boomerang lands near him, evidently belonging to a stranger *B* on the other bank who finds *A*'s spear instead. Somehow, with the aid of certain physical gestures, *A* leads *B* to believe that his faulty boomerang will be thrown back if he returns *A*'s spear first. *B* obliges but *A* leaves the boomerang where it is and continues on his hunt. Here, in the absence of any shared social practice or agreement that might assign obligations in advance, all that is conveyed is a conditional intention to perform an act. However, given that *A*'s expressed intention is to cause *B* to expect that act be performed, then *A* must be obliged to fulfil that expectation. In this case, it cannot be said that *B*'s act of returning *A*'s spear is sufficient to oblige *A* to respond alike. Such would be possible only under the presupposition of a social union with agreed rules for determining reciprocal obligations. To form a justified expectation of reciprocity, neither party need have any recourse to a prior agreement or custom. On the contrary, it is only from the mutual recognition of what we owe to each other that such agreements can evolve. Thus, *B* has a right to require that *A* return his boomerang, regardless of whether or not *A* has promised to do so. It is not a promise *simpliciter* that creates the debt. The practice of promising serves to formalise or publicise the debt so that it has greater weight in assigning responsibility. However, as this example shows, it is the expression of *A*'s *ability to respond* to the expectations given to *B* that obliges *A* to fulfil those expectations.

By signalling his intention to respond in kind, *A* induces *B* to expect that response, in the same way that any other belief is induced by the imputation of cause and effect. The concept of causal responsibility reflects the pragmatism of inductive reasoning that allows us to manipulate the environment to our own ends. Once we have developed sufficient awareness of our power as causal agents, it is naturally unavoidable that we should apply that same pragmatism to infer moral responsibilities. In our daily lives, we rely on our ability to foresee and induce changes in objects and events within reasonable parameters. If I am catching a train, I should reasonably expect the train to arrive and depart at a time closely approximating that advertised in the timetable. But precisely because the train's performance is under the control of other autonomous human beings, I cannot expect any of them to assume absolute responsibility to follow strict schedules.

Nevertheless, while the act of purchasing a valid ticket entitles me to expect that the advertised service will be received, that act only obliges the service provider because it satisfies a prior condition set by that provider. If the latter has not given any prior indication that a specific service will be given, it can be under no obligation to supply it. Where the timetable indicates that services will cease after midnight, I have no more reason to expect a train to appear at 2 a.m. than I do to expect that the sun will also appear at that time. Thus, whether within or without a social union, a person cannot simply acquire an obligation as a result of another's act. Otherwise, rights could be claimed by *fiat*. In the state of nature example, the spear thrower could simply assume the right to keep the boomerang for himself. Likewise, everyone would have a right to hi-jack a train to work everyday. Despite the initial improvement in frequency, such anarchic self-service would very soon produce no reliable service except for those who succeed in acting upon another assumed right to take control of the system. Of course, under such conditions, any expectation of service is unwarranted until the resumption of normal services is declared and their conditions and details published. Only then does my ticket become valid in acknowledging my acceptance of those conditions being fulfilled, which then gives me the justified

expectation that the advertised service will be provided in return. The rights and obligations of both parties are thereby acquired as a means of insuring the value of their mutual expectations. But the reasonable value of those expectations need not be predetermined by social conventions that might not represent the best interests of all parties. In recognition of this fact, 'normal' train services should be seen as the *result* of planning for the optimal value that can be delivered, without creating too many inconveniences or costs to either the customer or provider. Although individual passengers do not have the power to tailor schedules to suit their personal routines, the frequency of services is designed in response to the travelling habits of users. In this sense, a public transport system is representative of all other social institutions that are founded on principles of democratic pragmatism. Thus, a common social practice like promise-keeping endures because of its value in justifying obligations to others that would not otherwise be acknowledged or fulfilled. But again, it is not the normative rules attached to making promises that obliges *A* to keep his promise to *B*. The rules themselves are a conventional form of conveying a prior understanding of indebtedness that is inherent in the pragmatism of human reasoning.

As this relation of indebtedness is aroused by the *reasonable* expectations that we intentionally induce in each other, the obligation to meet such expectations is determined by the essential principle of practical reason – causation. It is *because A* has induced *B* to rightfully believe that *X* will be done that *A* is responsible for justifying *B*'s belief. When *A* subsequently fails to do *X*, he effectively falsifies *B*'s belief, thereby undermining the reliability of another's expressed intentions. Being unable to depend upon another's actual ability to respond would lead to a skepticism and solipsism that would be disastrous for any social animal. Dependent upon each other for survival, actual responsibility is a demonstrable imperative in the livelihood of all human beings. In this sense, duties to others can be defended as a natural requirement of human interdependence that does not require contractual agreement.

In fact, without the prior recognition of such duties, we would not even qualify as moral agents capable of judging how we might best govern ourselves in the light of such knowledge. Therefore, these mutually evolving responsibilities among autonomous agents must be factored into the process of determining which rights and obligations which can be reasonably accepted as universal moral constraints upon the actions of such agents. Accordingly, the *evolving autonomy* on which moral agents depend cannot be put behind a 'veil of ignorance', as it is this very capacity which *enables* them to make rational and reasonable choices of the principles and norms which can best protect and maintain that essential adaptive capacity. This chapter's broad discussion of mutually evolving social responsibilities will thus serve to inform the subsequent *organic* approach toward reconstructing the pragmatic intelligibility of a social contract to be presented in Part Three. However, as the organic development of practical reason and autonomy is crucial to each agent's ability to *reflectively control* their actions as a *moral* agent, the validity of normative standards of moral reasoning must first be assessed in order to avoid endorsing contractual principles and norms that cannot be rationally justified. The next chapter will therefore address the standard objections raised against naturalistic defences of moral reasoning.

[117] Aristotle, *Parts of Animals*, in *The Complete Works of Aristotle,* Revised Oxford Translation, ed. J. Barnes (Princeton, Princeton University Press, 1984). See II & III, esp. 662b.

[118] Hume, *A Treatise on Human Nature*, II. III. III, p.415

[119] Immanuel Kant, *Groundwork of the Metaphysic of Morals*, in *The Moral Law*, H. J. Paton trans. (London, Hutchinson University Library) 390.

[120] Ibid., 403

[121] Russell Hardin, *One for All: The Logic of Group Conflict*. (Princeton, Princeton University Press, 1995), p.205

[122] Ridley, op.cit., pp.191-192

[123] Philippa Foot, "Morality as a System of Hypothetical Imperatives," in *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford, Basil Blackwell, 1978), p.162

[124] Ibid., p.163

[125] Ibid., p.160 ff.

[126] Ibid., pp.162-163

[127] Ibid., pp.163-164

[128] See note 5 above.

[129] Note Hume's nonchalant remark: "'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger" in A *Treatise on Human Nature*, II. III. III p.416

[130] John Dewey, *Art as Experience* (New York, Capricorn Books, 1958), p.47

[131] Ibid., pp.26-29

[132] e.g. *Poetry, language, thought*, trans. A. Hofstadter (New York, Harper & Row, 1971). *The Piety of Thinking: Essays*, trans. J. G. Hart & J. C. Maraldo (Bloomington, Indiana University Press, 1976) and *Pathmarks*, ed. W. McNeill (New York, Cambridge University Press, 1998).

[133] Nietzsche, *On the Genealogy of Morals*, in *Basic Writings of Nietzsche*, p.596. See editor's note 1.

[134] See Maynard Smith, *Did Darwin Get It Right?*, p.188, for discussion of the possible significance of these random changes which can sometimes be "contrary to selection"

[135] Sterelny, *The Evolution of Agency*, pp.173-174

[136] See examples cited in Chapter 2.

[137] De Waal, *Good Natured*, pp.144-145

[138] Richard Alexander, *The Biology of Moral Systems* (New York, Aldine De Gruyter, 1987), pp.114-115

[139] James Lennox, "Aristotle on the Biological Roots of Virtue: The Natural History of Natural Virtue," in *Biology and the Foundation of Ethics*, ed. J. Maienschein & M. Ruse (Cambridge, Cambridge University Press, 1999). p.13

[140] Ibid., p.15

[141] Jean-Jacques Rousseau, *A Discourse on Inequality*. trans. M. W. Cranston (Harmondsworth, Penguin, 1984), pp.134-137

[142] Hume, A *Treatise on Human Nature*, III. II. I

[143] Lennox, op. cit., p.22

[144] Lennox, op. cit., p.24

[145] Aristotle, *The Nicomachean Ethics*, trans. H. Rackham (London, Wordsworth, 1996), II. VIII. 4-6

[146] Re planning of the Hiroshima bombing, Truman notes in his diaries that "I have told the Sec. Of War, Mr Stimson, to use it so that military objectives and soldiers are the target and not women and children. Even if the Japs are savages, ruthless, merciless and fanatic, we as the leader of the world for the common welfare cannot drop this terrible bomb on the old capital or the new….The target will be a purely military one and we will issue a warning statement asking the Japs to surrender and save lives." See Harry S. Truman, *Off The Record: The Private Papers of Harry S. Truman*, ed. R. H. Ferrell (New York, Harper & Row, 1980), pp.55-56; See Jonathan Glover, *Humanity: A Moral History of the Twentieth Century* (New Haven, CT, Yale University Press, 2000), pp.101ff. for criticism of Truman and others on the grounds of their moral "evasion" and "fragmentation of responsibility". In his essay, "Harry S. Truman, 1945-1953," in *Character Above All: Ten Presidents from FDR to George Bush*, ed. R. A. Wilson (New York; London, Simon & Schuster, 1995), David McCullough asks: "Now suppose you are the President and you know these horrible raids you are sending against Japan are killing as much as twice or three times that number, and that possibly one bomb delivered by one plane might, if it worked, have the effect of shocking the enemy into surrendering and all the slaughter could stop. An invasion of Japan would become unnecessary. Thousands of American lives would be saved. And imagine also that no one among your top advisors is telling you not to use the bomb. The President – the commander-in-chief – gave the order. The war ended."

[147] Allan Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge, Mass., Harvard University Press, 1990), p.26

[148] John Rawls, *A Theory of Justice*, pp.136-142

[149] Desiderius Erasmus, trans. M. M. Phillips *Adages*. (Toronto, University of Toronto Press, 1974), bk 3, century 4, no.96

[150] Hume, *A Treatise on Human Nature*, III. II. VI, p.526

[151] Hume, *A Treatise on Human Nature*, III. II. II, p.489

[152] Gibbard, op. cit., pp.72ff. & 263-264

[153] Wright, op. cit., p.143

[154] Marc Bloch, *Feudal Society*, Vol. 1, trans. L. A. Manyon (London, Routledge & Kegan Paul, 1962), pp.131ff.

[155] Hume, *A Treatise on Human Nature*, III. II. II, p.490

[156] J. L. Mackie, *Hume's Moral Theory* (London ; Boston, Routledge & K. Paul, 1980), p.89

[157] Ibid., p.90

[158] See Coveney & Highfield, op. cit., pp.230-2; Robert Axelrod, *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration* (Princeton, Princeton University Press, 1997), Chs.1 & 2

[159] Robert Axelrod, *The Evolution of Cooperation* (New York, Basic Books, 1984), pp.27-54

[160] Brian Skyrms, *Evolution of The Social Contract* (Cambridge, Cambridge University Press, 1996), p.60; See Axelrod, *The Complexity of Cooperation*, chs. 1 & 2 respectively, for discussion of richer models of ecological dynamics & realism of "noise" i.e. random errors in choices

[161] Skyrms, op.cit., pp.61-62

[162] Karl Marx et al, *Critique of the Gotha Programme* (New York, International Publishers, 1966), I. § 3, p.10

[163] Rawls, *A Theory of Justice*, p.78

[164] Though Kierkegaard is himself playing the devil's advocate here, by presenting his analysis in the imaginary form of an essay found among the papers of a young aesthete (*Either*), which is later critiqued in the imagined monologue of the elder "Judge William" who espouses the superiority of marriage and the ethical mode of life (*Or*). See Søren Kierkegaard, *Either/Or*. Vol. 1, trans. E. H. Hong, et al. (Princeton, Princeton University Press, 1987), p.3ff

[165] Ibid., p.93

[166] Ibid., p.94

[167] Harry G. Frankfurt, "Freedom of the will and the concept of a person," in *The Importance of What We Care About: Philosophical Essays* (Cambridge, Cambridge University Press, 1988), pp.16-17

[168] Kierkegaard, op. cit., vol. 1, p.97

[169] cf. possible interpretations of DJ as embodying the principle of "free exchange" which he ironically both subverts and upholds in capitalising on other's desires: see James Mandrell, *Don Juan and the Point of Honor: Seduction, Patriarchal Society, and Literary Tradition* (University Park, Pennsylvania; Pennsylvania State University Press, 1992), pp.252-256

[170] *Contra* Kierkegaard's dismissal of the "ludicrous" fact that Don Juan's servant, Leporello should keep a list of his master's conquests e.g. "1,003 in Spain" (See op. cit., vol. 1 p.95), Camus claims that while DJ *is* a seducer, he is not an immoralist given that "what Don Juan realizes in action is an ethic of quantity", which is certainly enough to make him an absurd man in an absurd world where "the value of a notion or of a life is measured by its sterility" - *The Myth of Sisyphus*, trans. J. O'Brien (Harmondsworth, Penguin Books, 1975, pp.69 & 66); cf. Tirso's original, 17th century version of DJ, who, in contrast to the "real and fictional libertines and sceptic figures" of the day, is portrayed as ultimately becoming a victim of *his* absolute trust in the hypocritical and untrustworthy behaviour of *others*, the myth thus still serving as a moral lesson in the social dangers of being "seduced by appearances": see Jeremy Robbins, "Seduced by Appearances: Seventeenth Century Sceptics, Libertines, and Tirso de Molina's Don Juan," in A. Ginger, J. Hobbs & H. Lewis (eds.) *Selected Interdisciplinary Essays on the Representation of the Don Juan Archetype in Myth and Culture* (Lewiston; Queenston; Lampeter; Edwin Mellen Press, 2000), p.36 & ff.

[171] Frankfurt, op. cit., p.14

[172] Kierkegaard, op. cit., Vol. 1, p.95

[173] This evolutionary dynamic in the myth is also reflected in Girard's anthropological interpretation of the "mimetic desire" which spreads thorough a community, inducing "envy, resentment and rivalry". The primitive 'moral' of such myths is evident in its focus upon the impending violence and threat to the social order, which can only be averted by the community uniting to direct its "collective antagonism" against a scapegoat who is then expelled or even killed: see Judith Hepler Arias "The Don Juan Myth: A Girardian Perspective," in D. Bevan ed. *Modern Myths* (Amsterdam & Atlanta GA; Rodopi, 1993), pp.24-26

[174] See Kierkegaard's rejection of Molière's comical version suggesting DJ's making marriage proposals to seduce – op. cit., vol. 1, pp.113-114

[175] For summation of these beginnings see Jean Piaget, "The First Year of Life of the Child," (1927) extract in *The Essential Piaget* (New York, Basic Books, 1977), pp.211-214

[176] Cf. Singer, *The Expanding Circle*, pp.157-159

[177] See Colin Turnbull, *The Mountain People* (London, Picador, 1974), pp.111-13: Threatened with starvation in an alien environment, Turnbull notes that, while their society survives, the Ik's familial bonds have fractured to the point where mothers happily abandon their children when they reach three years of age. But note Singer's remarks in *The Expanding Circle*, pp.25-6, on Turnbull exaggerating the extent of their ethical failings.

[178] Norman Daniels, *Reading Rawls: Critical Studies on Rawls' A Theory of Justice* (New York, Basic Books, 1975), pp.158-159

[179] Rawls, *A Theory of Justice*, p.19

[180] John Rawls, *Political Liberalism* (New York ; Chichester, Columbia University Press, 1996), pp.10-11

[181] Rawls, *A Theory of Justice*, p.99

[182] Ibid., p.462

[183] Thomas Scanlon, *What We Owe To Each Other* (Cambridge, Mass., Harvard University Press, 1998), pp.296-297

CHAPTER FOUR

CHAPTER FOUR

SELECTION AND REFLECTION:
PRACTICAL REASON AS AN ORGANIC PROCESS

(i)   The Pragmatic Limitations of 'Is/Ought' and 'Fact/Value' Distinctions

To claim that certain duties are naturally required as a result of the vital interdependence of human activities is liable to be seen as yet another impossible attempt to bridge the logical gap between what "is" and what "ought" to be. Even worse, it might be seized upon as the same paradigm case of the "naturalistic fallacy" that Moore had cited against Spencer's attempt to derive ethical conclusions from our evolutionary success as a species.[184] The veto against deriving values from facts has assumed something of a dogmatic status among contemporary analytical philosophers, to the point where even to entertain factual considerations as pertinent to moral ones is likely to provoke a scornful protest or pity, on the assumption that it entails a 'naïve' realism that is clearly untenable. However, if philosophers are to remain true to their task of questioning assumptions, they especially ought to resist falling into the kind of "dogmatic slumbers" that only obstructs further enquiry. Accordingly, the "is/ought", "fact/value" gap cannot be dismissed as axiomatic or necessary simply because the rudimentary rules of logic claim to demonstrate its truth. As Hare has convincingly shown, language exhibits a logic of descriptive entailment that is often inherently prescriptive at the same time.[185] Thus, in spite of the logical gap between fact and value in abstract, or 'pure' reason, the *teleological* nature of *practical* reason requires us to evaluate facts i.e objects and events, as means and ends for our projected activities.

The abstract distinction between description and evaluation is itself derived from another traditional philosophical distinction that has nevertheless been thoroughly debunked as a unfounded dogma.[186] In Putnam's words, the "history of the fact/value dichotomy parallels, in certain respects the history of the analytic/synthetic dichotomy" in that both derive from Hume's assumption of a "metaphysical dichotomy" between "matters of fact" and "relations of ideas".[187] In both cases, facts are distinguished from values or ideas on the basis of a 'pictorial' and associationist theory of mind which holds that ideas are constituted by sensory 'impressions' which 'resemble' their objects.[188] As the theory is unable to locate the empirical source of those kinds of ideas which do *not* resemble observable entities, the *meaning* of such ideas could only be described in terms of their abstract and tautological *relations*, as subsequently maintained in the analytic conception of the synonym of terms

like 'bachelor' and 'unmarried'. On this semantic basis, the analytic/synthetic distinction classifies propositions in terms of necessity and contingency. Thus, "All bachelors are unmarried" is held to be necessarily true because analysis of the meanings of subject and predicate suffices to determine the truth of the statement. By contrast, the truth of a synthetic proposition such as "Alan Gewirth is an American philosopher" is a contingent fact that can only be established by empirical evidence. However, as Quine has convincingly argued, the analyticity of a statement is itself dependent upon a prior understanding of the meanings of its terms.[189] The knowledge that bachelors must necessarily be unmarried is no less contingently determinable by experience than is the fact of Gewirth's identity as a philosopher. Even if necessary truth is restricted to purely tautological statements like 1+1=2, it remains true only because it purports to represent a law of relations between objects that has been established by prior experience of the world.

Therefore, the bifurcation of knowledge into items that represent "relations of ideas" as distinct from "matters of fact", as Hume describes it,[190] is unwarranted. Of course, the principles of abstract reasoning, as circumscribed by deductive logic or mathematics, do represent relations among ideas. But the truth value of any such relation cannot be ascertained unless the ideas themselves already have a normative meaning drawn from practical experience of the world. The 'abstract' proposition that 1+1=2 is necessarily true only because it represents an existential relation between objects of experience, which is itself only a contingent fact. Thus, necessary 'analytic' truth is ultimately dependent upon contingent 'synthetic' facts that are interpreted as having some prior normative significance attached to them. It is the *practical* value of knowing that 1+1=2 that defines its necessity.

As Hume noted, the very idea of necessity is itself *necessary* for both theoretical and practical knowledge. The ability to control the actions of objects and events entails an understanding of causation that is premissed upon the imputation of a "necessary connection" between them. Desired results can thereby be induced by employing appropriate and effective means as circumstances require. But this practical knowledge is inseparable from the understanding of causation that informs it. This awareness need not be explicit or scientific in its grasp, as evidenced by the chimpanzees' primitive tool use. But where the behaviour clearly implies an imaginative ability to foresee the likely results of a novel action without any prior experience of it, some intuition of necessity must be present, even if it is only a dim apprehension. Only from such beginnings can the roots of a fully theoretical reasoning emerge, as self-awareness and intelligence evolve to meet the practical necessities of life. Mathematics begins as the formulation of rules and symbols as a means of evaluating and recording the quantities of particular objects that can be sourced to meet those needs most efficiently. The artificial distinction between 'pure' and 'applied' branches of mathematics and other disciplines only maintains this dogmatic dualism of theory and practice. As such, it is not the validity of drawing such distinctions which is problematic, as many do indeed have a legitimate disciplinary role in structuring the flow of information and order of inferential reasoning in all areas of thought. Thus, while it is certainly useful to distinguish theory and practice in both scientific and moral reasoning, in philosophy the distinction reaches an apotheosis of absurdity in designating a realm of ethics as 'applied', as if there were a branch of ethics unconcerned with the practice of living. Perhaps aware of this absurdity, theories of ethics are not called pure, but "meta-ethics". Nevertheless, the meta-ethical enquiry into what is meant

by concepts of "good" or "right" is a misguided attempt to describe concepts of value in isolation from the factual conditions which give rise to them. As John Dewey notes, "If our moral judgments were just judgments *about* morality, this might be of scientific worth, but would lack moral significance, moral helpfulness. But moral judgements are judgements of ways to act, of deeds to do, of habits to form, of ends to cultivate."[191]

In the same way that the analytic/synthetic distinction attempts to isolate meaning from its roots in experience, the fact/value distinction pretends that judgements of value are somehow independent of the factual world from which they emerge and to which they must adapt, as conditions for the development and experience of new possibilities. More than fifty years before Quine's critique, Dewey raised a similar criticism against the empiricist conception of ideas as induced by "impressions", so that "the idea arises as a reflex of some existing object or fact. Hence the test of its objectivity is the faithfulness with which it reproduces that object as a copy."[192] By contrast, Dewey defends an historical, or 'genetic' account which "holds that the idea arises a response, and that the test of its validity is found in its later career as manifested with reference to the needs of the situation that evoked it."[193] Thus the traditional empiricist account is fundamentally misconceived in that it fails to recognise that ideas do not serve merely to reflect reality, but to transcend or reconstitute it for some purpose. As Dewey observes:[194]

> The idea of withdrawing the hand may be an adequate response to the perception of a flame. The idea, however, is not an impression of the object. In like manner the notion of giving an accused man a chance to justify himself may be an adequate response to the stimulus of capture and presumed guilt. And yet it in no way depends for its reality upon being a mere impress of the existing state of affairs. The test of its worth is its capacity to regulate the various factors entering into the situation.

It is in this sense that is/ought, fact/value distinctions parallel the analytic/synthetic distinction, in that all such distinctions are arguably based on a now debunked Lockean conception of the mind as *tabula rasa*. In Locke's theory, ideas are identified in terms of a purported distinction between nominal and real essences. While the latter 'ectypes' represent copies or images of real 'substances', practical and moral ideas have a nominal essence in being 'archetypes' constructed not for the purpose of representing real substances, but rather to conceptualise *possibilities yet to be realised*.[195] As such, Locke's metaphysical distinction corresponds to a genuine distinction that a number of contemporary philosophers have described in terms of a concept's "direction of fit".[196] This critical distinction echoes Dewey's observations above and is neatly described in the following quotation from Platts:[197]

> The distinction is in terms of the *direction of fit* of mental states to the world. Beliefs aim at being true, and their being true is their fitting the world; falsity is a decisive failing in a belief, and false beliefs should be discarded; beliefs should be changed to fit the world, not vice versa. Desires aim at realization, and their realization is the world fitting with them; the fact that the indicative content of a desire is not realised in the world is not yet a failing *in the desire*, and not yet any reason to discard the desire; the world, crudely, should be changed to fit with our desires, not vice versa.

It can now be seen how this legitimate distinction in direction of fit reflects the is/ought, fact/value distinctions, as each essentially highlights a genuine distinction between ideas which function to represent real objects or events, as opposed to those which serve to guide action toward the realisation of desired ends. However, as Putman argues, when the distinction is inflated into an unbridgeable metaphysical dichotomy or dualism where practical possibilities are held to be *independent* of facts, then it assumes the dogmatic status of a *petitio principii*.[198] Yet, as shown by Quine's critique of the analytic/synthetic distinction, there is no epistemological basis for insisting that the criteria used to determine the validity of a logical proposition is somehow not subject to the same direction of fit as all other propositions. In effect, the above quotation itself shows how the distinction between the direction of fit in beliefs and desires (as with facts and values) is exaggerated so that the latter appear to be able to persist without themselves having to fit with the former. However, Humberstone implicitly recognises the significant role of *both* directions of fit in evolutionary terms, noting that "in the case of the mechanism for desire, what is required is that the way the world is affected by the organism be causally dependent on the desires possessed."[199] In effect, by acting to realise their desires, individuals remodel the world so that those desires *can* be realised. But precisely because 'ought' implies 'can', certain desires logically *ought* to be discarded if they are *impossible to realise*. Thus, while a person may wish to be younger or to be someone else, as those desires cannot be realised *given the facts*, it is perfectly legitimate to claim that those desires ought to be abandoned for precisely the same reason that false beliefs ought be abandoned for misrepresenting the world. Note that such inferences are justified not simply by virtue of certain desires being supervenient upon certain facts, as in Hare's example of the impossibility in claiming that some painting 'P is exactly like Q in all respects save this one, that P is a good picture and Q not."[200] While those cases are constrained by logical consistency in the attributes being valued, desires or values which are *necessarily* unrealisable are directly inconsistent with the laws governing the properties and actions of physical phenomena. Thus, while strictly deductive logical necessity imposes constraints on the coherence *among* facts and values, there are demonstrable *synthetic* relations of *nomic* necessity between certain facts and values which unproblematically warrant *inductive* inferences that some beliefs ought to be abandoned. Thus, such inferential criteria go beyond questions of mere sensitivity to conventional facts or physical difficulties, like whether or not one ought to aspire to become President of the United States in spite of being born in Australia, or even to climb Mt. Everest in spite of some disability. Though such facts do represent obstacles to the successful realisation of certain desires, they do not contravene physical laws as the belief that one can become younger or someone else does.

Similarly, as Rescher argues, the alethic criteria which justifies the cognitively inductive inference from "There looks to be a cat on the mat" to the positive belief that "There is a cat on the mat" proceeds from "essentially the same standard" of "cogent systematisation" of *experience* which also belongs to the criteria justifying axiological judgements.[201] Whether or not one *ought* to believe "there is a cat on the mat" also involves "the problem of bridging a (seemingly insuperable) gap between appearance and reality, between phenomenally subjective claims at the level of appearances and impressions, and ontologically objective claims at the level of being and actuality."[202] Thus, as will be shown in Dewey's theory below, inductive reasoning is a process which acknowledges an

indeterminacy and fallibility in the causal connections perceived in phenomena which applies to both *factual* beliefs and judgements of the probable *value* of conceivable means and ends. Ironically, this epistemic gap is problematic because of the 'problem' of induction noted by Hume himself, and it is arguably Hume's own recognition of this *inductive* gap which underlies his criticism of 'every system of morality' hitherto encountered, as these cannot circumvent this problem by attempting, as MacIntyre suggests, "to render inductive arguments deductive…".[203] However, as Hume also proceeds to show how the rules of justice *can* be inductively justified in terms of common social interests and needs, Hume himself provides a sound pragmatic criterion for bridging the is/ought gap.[204] Given that these *psychological facts* are also practical *ideas*, they serve as synthetic bridge principles undermining the validity of an epistemic *dichotomy* between "relations of ideas"/"matters of fact", for precisely the same reason that the "borderline between almost analytic and synthetic is rather vague and permeable."[205] Likewise, as shown above, the direction of 'fit' between beliefs aimed at *representing reality* and that which desires aim to *realize* must also converge at some point if beliefs and desires are to be effective guides to action.

Ultimately, the fundamental error in the fact/value dichotomy consists in reducing the validity of a belief to a measure of its correspondence to some *existing* facts, rather than as an *evaluation* of those facts as a condition for *further* analysis and action. As Quine has clearly demonstrated, without any prior synthesis of experience, the analysis of related meanings cannot even begin. The hypothesis that 1+1=2 is determined by the fact that it works in practice. As judgements of value are prescriptions for determining and controlling social and environmental conditions, they are themselves continually subject to the constraints of prevailing conditions. Therefore, the normative question of what *is* good or right is only answerable in the light of those facts which can serve as objective means of bringing about the valued end. By imposing limits upon the feasibility of means and ends, the environment conditions the range of values or interests that can find sufficient reward to sustain the organism's motivation to act.

This pragmatic evaluation of means and ends in relation to facts is a process of progressively testing ideas for their value in guiding the organism towards the resolution of a problem. Dewey recognised that such a dialectical process is most accurately represented as it is experienced by the individual who perceives the situation as problematic. Only then can the exigency of factual considerations be properly grasped and assessed. In analysing the process of reflective thought, Dewey highlights five stages or functions whereby the situation is progressively enlightened and resolved.[206] These will be examined in more detail in section (iii) of this chapter. However, as this evolutionary process plays a central role in the next section (ii) on the dialectical ontogeny of moral learning, it will be useful to provide a preliminary sketch of Dewey's model by his own example. The reader is asked to imagine a simple case where you are walking somewhere without any discernible path,[207] when you suddenly find a ditch blocking your way. The fact of being confronted by this obstacle provokes the conception of a initial plan or suggestion – in this case, the act of jumping the ditch. But if the plan is to succeed, it must be checked against the actual conditions which reveal the ditch to be quite wide, while the opposite bank appears slippery. Those observed hindrances lead you to imagine an alternative idea. Somewhere else the ditch may in fact be narrow enough to jump. But again, observations show this not

to be the case. Surveying the area for other possibilities, you see a log which might used as a bridge. Judging by its length and width, it appears adequate for the task and should not be too heavy to move. This evaluation is then tested by executing the plan. You succeed in putting the log across the ditch and crossing it, thereby confirming the value of the idea.[208]

This simple illustration shows how facts are indispensable determinants of objective meaning and value. However, the mere presence of the log is not in itself indicative of any definite value for an observer who is not in a similarly problematic situation For a bushwalker on the look out for a campsite, the log presents quite different possibilities. If it is near dusk and there is a chill in the air, the log will likely be regarded as a potential solution to the problem of staying warm. But whatever the case, its value is derived, not from the log itself, but from the effect which it may produce. The log's value as a means is determined by the value of the end. In support of the fact/value distinction, it is argued that ends are only possibilities for action that may be accepted or rejected. Even if few would rationally choose the destruction of the universe, Dewey agrees with Hume that, in deliberating about ends, "there is nothing in the wide universe which may not…be accepted or rejected".[209] However, the *fact* that the desirability of ends need only be limited by the imagination only confirms the *value* of the imagination. Thus, even though value judgments transcend factual considerations, such judgments depend upon the *actual* ability to infer the instrumental efficacy afforded by various objects and events.

To conceive and evaluate a philosophical hypothesis, such as the is/ought distinction, requires that one recognise the value of reason itself as a means of discovering uniform relations among phenomena in experience. Thus, the criterion of reasonability also applies to the is/ought distinction *itself*. In fact, the distinction itself *is* an is/ought inference. For making any theoretical distinction is also an *act,* an evaluative judgment that it *ought* to apply. If we can find no good *pragmatic* reason that it should, then it has no warrant in practice. The validity of such a distinction cannot be asserted or maintained solely on the basis of its conformity to certain logical laws, as if human psychology and behaviour must be required to fit such strict patterns. On the contrary, as the laws of inferential reasoning are revealed through the power of human agency in the world, the cognitive process of evaluation cannot be represented as a purely mental event divorced from the actual conditions that precipitate it.[210] If the various meanings assigned to facts had no particular relevance to human interests, it would indeed be a mystery as to how they obtained such meanings. On this basis the subjectivist is justified in concluding that all meanings are products of the mind given to phenomena. However, it does not follow that values, desires and even everyday beliefs are arbitrarily imposed upon items of interest, before they are experienced as such. This extreme kind of global subjectivism is seriously flawed in its failure to explain how interest in any object can even be aroused in the first place, unless it is directed by some aspect of the environment in which it is situated.[211]

In the above example, it was shown that the value of the log is, of course, subjectively determined by the interests of the person concerned. It is only my desire to traverse the ditch that leads me to perceive the log as a means to that end. If I were tired and less curious, I may well prefer to end my excursion at the ditch and set up camp for the night. In that case, I will be inclined to appraise the same log as potential firewood. Whatever the circumstances, neither the bare existence of the log, nor any inherent quality is responsible for defining its value. It is the fundamental indeterminacy of value

judgements that Moore cites against the possibility of defining "good" in terms of any given quality. However, by the same token, Moore recognised that, although it is always possible to ask whether a given quality is good, that quality must have some objective reference. Though it is indeterminate, revisable and thus "unanalysable" as a general term, in a specific context the use of the term "good" must entail a description of something, if it is to convey *any* meaning. Otherwise, no-one would ever be able to understand, agree or disagree with another's judgement of goodness. Likewise, the revisability of value judgements should not be taken as an indication of their lacking objectivity. In contrast to *desires*, values and ideals are not the result of cultivating some arbitrary whim. As considered judgments, they are informed and influenced by the personal and collective experiences accumulated in the social environment. From this wealth of prior experience, each individual derives their own unique range of personal values which are themselves continuously reevaluated in the light of their actual effects. However, as the essential evolutionary dimensions of this process begin in early childhood, research in the psychology of moral development provides an invaluable source of evidence for the validity of the Deweyan model, which will then be re-examined in the light of this research.

(ii)   The Dialectical Ontogeny of Moral Learning

Theories of moral development have been dominated by Piaget's experimental research revealing distinctive stages in the ontogenetic process whereby the child progressively learns to apply general principles of reasoning to interpreting and controlling its social environment. As such, moral judgement is essentially guided by the same intellectual operations through which the behaviour of physical phenomena is grasped i.e. via an 'equilibration' defined as "an equilibrium between the structures of the subject and the objects; [the subject's] structures accommodate to the new object being presented and the object is assimilated into the structures…"[212]. In relation to individual development, this "equilibrium factor" is manifested in "exchanges between the organism and the environment" which have a fundamentally adaptive function which Piaget has summarised thus:[213]

> All adaptation, both mental and physical, includes two poles: one corresponding to the assimilation of energy or matter from the environment by the structure of the organism (or mental assimilation of data perceived in the environment to the schemata of action followed by the subject); the other corresponding to the accommodation of structures of schemata of the organism or subject to environmental situations or data. Adaptation is then nothing more than an equilibrium between this assimilation and organic or mental accommodation. This is why the most elementary exchanges between subject and object are already determined by the equilibrium factor.

> Piaget's theory is clearly indebted to Baldwin's initial formulations of ontogenetic stages and their role in the organic selection process operative in phylogenetic evolution.[214] As Kohlberg's subsequent six-stage theory represents a more discriminating analysis and extension of Piaget's three basic cognitive stages in moral development, it is based on empirical findings which only serve to lend further support to Piaget's original research. Nevertheless, both Piaget's and Kohlberg's theories have been subject to criticisms on the basis of more recent discoveries supporting the recognition of

'domain-specificity' in both children's and adults' reasoning about moral and social issues.[215] Both Piaget's and Kohlberg's stages characterise moral reasoning as essentially exhibited in distinct shifts in the individual's social orientation. Obedience to authorities is seen as initially motivated by rewards and punishments which later come to be appreciated in terms of the value of social conventions, established rights and duties and finally in a conception of the universality of justice and equality. [216] In effect, the stages encapsulate the progression from heteronomy to autonomy in moral judgement, as a reflection of the child's own growing consciousness of the inherently rational grounds for justifying the validity of different moral principles and norms. However, this natural growth in the child's powers of reasoning led Kohlberg to conceive the sequence of stages as both global and invariant, for which he was also criticised, particularly as his research was originally based on the responses of seventy five American youths presented with a number of hypothetical moral dilemmas, "some found in medieval works of casuistry."[217] While these stages have since been shown to be less rigid in sequence and also fail to capture other salient and variable aspects of moral judgement, Kohlberg's model has largely withstood criticisms of cultural and gender bias.[218]

By contrast, the domain-specific research has shown strong evidence indicating that both children and adults of all ages are in fact capable of acknowledging and differentiating between the modes of reasoning corresponding to two or more of Kohlberg's stages. These studies show that children typically distinguish between three co-existing conceptual domains: a moral domain which encompasses concepts of "the welfare and rights of persons" which includes rules such as those "against hitting and stealing"; a domain recognising rules of social conventions such as those governing "forms of address and table manners"; the third domain is associated with the "psychological characteristics of persons, such as their beliefs, ideas and feelings", part of which involves the respect for personal autonomy in issues such as the "choice of friends or style of dress".[219] Numerous examples from studies of peer interactions among toddlers and preschoolers reveal a clear moral sense in their responses to acts involving "physical and psychological harm, fair distribution, and the violation of rights". Thus, conflicts involving "aggression (hitting and hurting), possession of objects (taking a toy or not sharing), teasing, and name calling" all invoke responses (explicitly defended by reasons like "She wouldn't let Agnes help", "You don't have to be rude" or "Mine! Mine!") which acknowledge an emerging conception of moral rights and obligations, equality and fairness.[220] Furthermore, even preschoolers have some regard for the domain of personal autonomy, in actively challenging and negotiating with their mothers on the extent to which they should perform a favour such a helping with the mother's cleaning up.[221]

The domain-specific research also shows that "even very young children do not view moral rules as sacred, unchangeable and contingent on the demands of adult authority", thus refuting both Kohlberg's and Piaget's depiction of the child's early conceptions as essentially heteronomous.[222] Instead, it is the domain of social conventions whose authority is judged as heteronomous, yet still acknowledged to be "relative to the social context, alterable, subject to authority jurisdiction, and contingent on rules."[223] Summarising this extensive research, Smetana notes that "rather than forming a developmental sequence, heteronomy, moral autonomy, and autonomy from moral and societal

constraint are found throughout development in children's judgments of social convention, morality, and personal issues, respectively."[224]

While these findings clearly undermine Kohlberg's insistence upon an invariant sequence in stages of moral development, they are not inconsistent with Piaget's account, as he was always more circumspect in admitting the possibility that such stages may instead represent "phases characterizing certain limited processes".[225] Indeed, Piaget's later theory of cognitive stages allows for an "asynchrony" or delay in stages of moral reasoning, precisely because these *follow* from and are *guided by* the child's developing intellectual skills.[226] Similarly, Piaget's theory can easily be reconciled with the evidence for domain-specificity, as it precisely the changes in the *social contexts* of the child's interactions which provokes the need to reconceptualise its social experiences in terms of a new set of functions. Thus, in moving from the experience of hierarchical, asymmetric interactions with parents and siblings into a context of egalitarian exchanges with peers, the child must begin to apply a new "working model" to *assimilate* and *accommodate* such new behaviour patterns into its repertoire of adaptive social skills.[227] For the same reason, even the distinct stages witnessed in Kohlberg's studies can be accounted for in terms of the child's judgement being guided by one or more moral *principles* which are applicable to a range of issues across one or more domains of social behaviour.[228] Thus, principles of equality and fairness, for example, both operate in evaluating the justice of social conventions, so that rules conferring unfair advantages are rejected. Likewise, in rejecting the authority of parents and teachers to override the child's personal autonomy in choosing friends, or style of dress etc., notions of potential inequality and injustice are implicitly inferred in the child's recognition of the unwarranted *disparities* between their liberties and those of their parents and teachers.[229]

Consistent with Baldwin's theory of organic selection, Piaget's theory of mental and moral development contrasts with Darwinian selection in being *assimilatory* rather than eliminative.[230] Thus, moral systems cannot be adequately construed as evolving solely by the selective deletion of those concepts, rules and beliefs which prove to have detrimental effects on well-being. This explains why all manner of beliefs or practices can *theoretically* be entertained, yet still regarded as *practically* inconceivable in terms of established moral principles. An obvious example is in circumstances of war, where a deeply entrenched moral principle against killing fellow human beings is temporarily and partially discarded. For the same reason, it is no accident that the most barbaric ideas and practices often re-emerge in wartime, as they are never entirely eliminated from the realm of possibility and are rationalised into action by the very process of assimilation and accommodation described by Piaget e.g. a once discredited idea like torture is revived and rationalised to fit with the 'new realities' of war and terrorism. In the absence of robust cultural institutions, the ontogeny of moral learning is therefore open to corruption and it is the formulation of strategies to *combat* this tendency which justifies the contractual approach that will be advanced in Chapters Eight and Nine.

Piaget's theory also accords with the 'direction of fit' analysis in confirming that "one can learn from experience either by exogenous perturbation or by anticipation."[231] In other words, a change in the environment pressures the organism to refit its concepts to represent those new conditions. But intentional, autonomous agents, by definition, are also self-motivating in their capacity to conceive new ideas to which the world may subsequently be fitted. Again, this account also accords with Baldwin's

early theory which was also formulated on the basis of his own observations of such processes in child development, where an environmental stimulus prompts "suggestions" for articulating fitting responses which then serve as ideas for guiding further adaptive actions.[232]

The mutual exchanges between organism and environment revealed in Piaget's findings can thus be seen as a dialectical movement between these two 'poles' or 'directions of fit', until a pragmatically effective equilibrium is reached in the realisation of some adaptively beneficial goal. By this process, moral principles are discovered and learnt through the *bidirectional* interactions which children and adults of all ages continually experience. This bidirectionality has been consistently observed in numerous studies of children's interactions with parents, peers and other caregivers, refuting earlier theoretical assumptions of a unidirectional process of 'internalisation' or 'socialisation', where the child is depicted as a largely passive product of the social environment. As Kuczinski et al. note, "Parents do not passively upload and download information from each other. Rather they act on their interactions and communications and interpret, select, forget, and reject ideas as they process or manifest the content of their working models."[233] This bidirectionality is also consistent with the domain-specific research, as it shows how the child learns to differentiate among the functions of various moral concepts, principles and practices by observing their activation in distinct social contexts. Even in the asymmetrical exchanges with parents, children exercise their own power of agency in learning to implement strategies such as issuing "commands" and "requests", negotiating over conflicts and evaluating the parents' actions. In the domain of conventions, as already noted in Piaget's early study where children themselves effectively "invent the social contract", recent research shows that "even preschoolers socialize each other into the skills of co-operative play and self-disclosure."[234]

All of the above research contributes to confirming an ontogenetic account of moral reasoning which has been well summarised in the following description by Hart and Killen:[235]

> The individual constructs moral principles by participating in social interactions, encountering social problems with moral dimensions (e.g., sharing, harming others, turn taking), formulating solutions to these problems through the use of the child's current moral principle adapted for the situation through either symbolic social exchange (role taking) or genuine collaboration with others (co-construction), and finally observing the satisfactoriness of the solution once it is implemented.

A significant virtue of much of this recent research is its emphasis on the importance of investigating "how moral judgments are constructed out of daily experiences."[236] Given that such judgements only evolve into *habits* and *norms* through regular, 'mundane' experiences and dilemmas (as opposed to 'runaway trains' and other rarefied 'possible world' events), it is in these commonplace scenarios that moral development is most reliably observed. Accordingly, elements described in the above processes can be more readily appreciated by considering examples of their operation at different stages and in different everyday contexts, beginning with the element of suggestion.

As each person can never be sure that an expected good is worthy of that estimation, it remains only a suggestion. Nevertheless, even the vaguest hint of what might be a good end does not spring spontaneously to mind in the form of an abstract concept. When reflective thought is prompted

into action by the apprehension of some object or event as a possible end to be realized, many of these immediate ends also operate as means for bringing about more long-term ends concerned with meeting the basic necessities of life. From the very beginning values are associated with the actual data of experience. A child who imagines it would be good to be a pop star or an athlete is not content with merely contemplating the possibility in daydreams. Curiosity demands that the idea be played out in practice immediately to experience the excitement of being a pop star or athlete. Thus, even in the young child's mind, ideas do not replace actions. On the contrary, the idea is explicitly conceived as involving a range of goals to be achieved.[237] In the act of pretending, the child imitates and rehearses the pursuit of these goals using whatever available means to represent the real activity, in the hope that the actual rewards to be experienced in these achievements will be worthwhile. If no such rewards were expected, there would be no cause for the child to imagine the occupation to be good in the first place. But in that case, the subjectivist will have to admit that such expectations are not spontaneous or contingent thoughts with no basis in fact. For such ideas are derived from the child's own observations and impressions, however naïve. Based on the facts as they are presented, the child has formed the belief that the activities of celebrities must be good, where good is generally identified with excitement. Here then, the meaning of good is not only associated with the performance of certain, definite acts that arouse excitement, but it has that value for the child due to the facts being presented as such.

As the child grows older, beliefs, values and even basic desires are continually formed by and tested against a background of prevailing facts. Gradually the child learns that certain physical laws of nature prevent the realization of events that were previously thought possible.[238] How can Santa Claus deliver billions of gifts to billions of children in one night? Such a question is prompted by the child's practical experience of the world which tests the validity of its beliefs. But this process of practical evaluation is not restricted to the determination of beliefs. Values and desires are subject to the same factual considerations that determine the degree to which beliefs are warranted. A change in the assessment of facts that renders a belief more or less justified, also calls for a reassessment of those ends which were initially valued or desired on the basis that they were *permitted by the facts* as they were then perceived.

When desires and appetites are confronted by stubborn facts that frustrate their satisfaction, those facts put pressure on the organism to adjust its response to those impulses. As shown earlier, moral values are typically the result of this environmental pressure to inhibit and direct instinctive desires towards socially productive ends that enhance survival and well-being. In this way, beliefs, desires and values evolve in direct response to factual conditions through the application of practical reason. But the process does not end there. While factual conditions induce the selection of values that can be usefully applied to objects and events, those conditions are altered in turn, as a result of the organism's changed behaviour. Finally convinced that the facts do not permit the existence of Santa Claus, the child recognises that the desire for gifts must be addressed to the *right* source i.e. parents. Now knowing the true facts of the matter, the child adapts its behaviour to them, so as to maximise the opportunities for receiving gifts. Santa's disappearance is scarcely lamented, when he is replaced by people who might even be begged and cajoled into buying Christmas presents everyday.

The child's insistent appeals constitute a new fact for the parents to deal with, in turn. Consequently, they too must reevaluate the situation to prevent the child from forming an even greater desire that is based on false expectations of others.[239] The parents may experiment with various rewards and punishments, in an attempt to discourage the child from making demands that cannot be fully met. Such a strategy may be necessary to protect the child from suffering the disappointments that will inevitably follow such 'great expectations'. But emotional conditioning alone does little to diminish the strength of the child's wishes.[240] Only by understanding why such wishes cannot come true, can the child become reconciled to the *fact* that those wishes are too strong in the circumstances. For instance, the certain prospect of being punished may effectively prevent a child from taking whatever it likes from the supermarket shelves. While the punishment is indeed a fact to be dealt with, it is explicitly presented to the child as a fact that need not eventuate, for its existence depends upon the intended desires of both child and parent. However, knowledge of the fact that some desires can and ought to be resisted, if only to avoid disappointment or punishment, does not yet give the child sufficient reason to evaluate those facts as good or bad. Furthermore, the subjectivist would argue that precisely because such possibilities only become facts if they are sufficiently desired, then that only confirms the priority of the passions in determining what ought to count as sufficient reasons for action.

But this objection misrepresents the psychology of desire and deliberation, by abstracting those very elements of the situation that issue in the demand for action. Firstly, it ignores the fact that ordinarily desires are not spontaneous eruptions that are somehow disconnected from the physical environment in which they arise. On the contrary, experience shows that desires are most often aroused by environmental stimuli, which accounts for their sudden appearance and fluctuations in strength and immediacy.[241] Secondly, the subjectivist cannot avoid acknowledging that desires are themselves *facts* that require an evaluation of the present and future conditions that are likely to affect their satisfaction. Both the appearance and subsequent evaluation of any desire depends upon the acknowledged salience of facts and the possibilities they permit in accommodating that desire's expression and fulfilment. However, while desires and values are causally dependent upon the awareness of relevant facts, including other desires and values as psychological facts, this does not mean that the former are necessarily determined by the latter. The selection of facts deemed to be relevant by any individual is the result of an innumerable combination of variable influences. Owing to these constant changes in perception and judgment, even the same individual in the same situation may thoroughly revise the importance ascribed to various elements. What was unnoticed only a moment ago may suddenly appear so vital as to reduce all else to insignificance.

It is a common cause of complaint and regret that decisions are frequently ill-informed or hasty. The persuasive urgency of desires that are left unexamined, or values that are held unquestioned, also contribute to the manner in which facts are selected and appraised in coming to any individual or group decision. These psychological elements are indeed subjective in the sense that they are unique to particular individual or group interests. But they also constitute genuinely objective phenomena manifested in the actions and reactions they sustain, as an organic feature of both the natural and social environment. As such, an accurate representation of what is meant by 'fact' and 'value' would have no basis for finding an unbridgeable gap between the two. Rather, subjective experience and objective fact

are inseparable constituents in the deliberative process. The beliefs, desires and values that a person brings to a new situation are already the product of a dialectical process of constant interaction with a variety of phenomena encountered in many other contexts.

In the case of the child in the supermarket, the desire to possess or consume some particular item is aroused into consciousness by the child's own perception and anticipation of the enjoyment it *ought to produce*. The use of the term 'ought' here reveals the *hypothetical* nature of the judgment and its immediate link to the desire that is provoked by the perceptual recognition of an object. As soon as the object is identified as 'Easter egg', its meaning is implicitly associated with the prospect of a certain pleasure, even if the child has not yet experienced that pleasure. Again, the enthusiastic approval of parents and other children will convince the child that these eggs *must* taste a lot better than those in shells, which are routinely spat out or thrown in disgust. Here, such unanimous expressions of enjoyment warrant an expectation of near certainty in the child's estimation of the egg's goodness. However, it is most important to note that these judgments are not the result of attributing goodness to the object in itself, unrelated to the interests of the human observer. Evaluations of an item's goodness are essentially gained through direct or indirect awareness of its possible value in producing some desired effect.

The judgment that "Easter eggs are good" only makes sense because it is understood as an abbreviated description of the fact that these products are generally *good to eat*, where the meaning of good is strictly understood as the enjoyment produced by the act of eating. By contrast, being urged to eat scrambled eggs for the first time, because they are "good for you" more likely to be taken as a sure sign that they *must not* be good to eat. Whatever the case, the meaning of the value judgment is specified in its application to various actions that function as means in the completion of some "end-in-view".[242] As such, the final conviction that something is good, marks the end of a process of evaluation that begins with the tentative belief that it might be good for some end-in-view. The specificity of the end-in-view determines the context and scope of activities which might be good for the realization of that end. In this sense, Moore's "open question" argument is misdirected in refuting the possibility that value can be ultimately identified with some definite quality. As a variety of possible ends are always conceivable, there will always be a corresponding variety of values to define the meaning of any given end. But the actual context in which those ends are conceived does define the feasibility of ends and means that are *in fact* likely to come to fruition or implementation.

For the mature agent, enquiry into the value of any phenomenon is never an entirely open question, as the questions that can yield practical answers are themselves the result of an earlier process of enquiry that has already succeeded in selecting the ideal range of productive activities that can be effectively realized in the circumstances. Concepts, beliefs and values are already defined within a specific context when they are grasped. Therefore, it is hardly surprising that 'good' should prove to be "unanalysable" when it is abstracted from the context which defines its scope of reference. As the term derives its meaning only by reference to some item in question, analysis of its meaning requires an analysis of the effects of that particular item in the context in which it is revealed.

Although, strictly speaking, every new situation is composed of a unique set of influences and experiences for each individual concerned, the past experiences of others in similar situations function

as cultural norms to guide the conduct of those who follow. With prior knowledge of the practical value afforded by various means and ends under various conditions, the question of what is good, or ought to be done in any given circumstances cannot be approached with naïve ignorance. To pretend that such questions cannot be definitively answered amounts to a denial of the possibility of ever taking effective action to achieve our goals. If factual knowledge did not have any prescriptive value for future actions, all decisions and plans would amount to nothing but wishful thinking from which nothing would be learnt. Of course, no amount of facts can *require* the performance of any act, any more than the possession of an overwhelming desire requires its satisfaction. As conditions for intelligent action, facts are therefore necessary but not sufficient for yielding a singular inference as to the 'right' response to a particular situation. In this abstract sense, values can never be logically reduced to any facts, as if such knowledge alone would be enough to yield the same evaluation on every occasion. For every value judgment is not only informed and directed by established norms. Those existing norms are continually subject to reevaluation in the light of the further explorations of present and future generations of individuals. As everyone brings a unique combination of personal endowments into the evaluative process, there can never be an ideal set of values that would be right for every person to apply in similar situations. Differences in personal histories and abilities, combined with differences in the elements defining a situation, will always ensure a relative indeterminacy of value. Nevertheless, this indeterminacy does not provide an argument for value being independent of fact. Given that these continual variations in person and situation *are* the facts necessary to produce corresponding changes in values, then the constitution of values is entirely dependent upon the constitution of those relevant facts, in the "supervenient" sense which Hare invokes.[243]

Behind the variations in individual and group interests is the selection process itself. Once it is recognised that beliefs, desires and values evolve in adaptation to the environment, it can no longer be argued that mental events are arbitrary or spontaneous expressions of subjective autonomy. Furthermore, the capacity for autonomous action is itself subject to environmental pressures, and conversely, that same capacity also enables the organism to initiate adaptive responses designed to selectively alter the environment to meet the organism's needs and interests more effectively. Therefore, the characteristic features of an organism's behaviour cannot be interpreted in isolation from its functional capacity in a given environment. Likewise, the environment cannot be adequately represented by ignoring the effects of the organism's behaviour in transforming it.

Although we have largely succeeded in reconstructing the environment to suit our collective ends, we are no less subject to the selection pressures imposed by these artificial conditions. An individual's choices are unavoidably constrained by the force of those conditions. The economic and social institutions that evolved to overcome problems of scarcity, now function as the forces to which the individual must adapt. However, as these regulative institutions are nothing more than the product of the co-ordinated, adaptive responses of individuals themselves, each participant is responsible for the critical evaluation and implementation of the norms that govern them. To a large extent, organism and environment function in symbiosis, the actions of the one provoking a reaction in the other and so on. Therefore, the nature and development of organism/environment, person/situation, subject/object,

fact/value is a dialectical process of mutual dependence. An analysis of either entity must proceed upon the recognition of the synthetic relationship by which each obtains its meaning from the other.

For the child in the supermarket, the perceptual recognition of a definite quality associated with an item is sufficient to stimulate a desire to experience that quality. The subjective desire to eat the Easter egg is aroused by noting the actual appearance of an object belonging to a class that has already been evaluated as 'good to eat'. In this case, the evaluation has been established on the basis of a consensus of experience transmitted to the child, so that anything made of chocolate provides cause for 'salivation', as it were, even before the child has experienced the taste. However, the strength of approval expressed by others remains hypothetical until the child actually 'tries' some personally. Sweetness is an objectively discernible quality whose universal appeal is easily explained by its evolved association with high energy food sources. Thus, it is a biological fact of no small importance that chocolate *is* good as an objectively valuable source of nutrition. Of course, it is generally not the nutritional value, but the refined sweetness that is enjoyed in the eating of chocolate. Furthermore, pure enjoyment does not constitute an evaluation, least of all when it is so strongly induced by biological factors beyond one's control. Yet, it is precisely such facts which must be taken into account in order to determine the value of eating chocolate on any particular occasion, as it is only upon knowing the likely effects of the act that any evaluation can be made. If any of those effects should prove to be contrary to expectations, or hinder the successful pursuit of other valued ends, these new found facts become the basis for reassessing the value of the activity in question.

As the consequences of the proposed activity are tried and tested by the individual or group within a certain context, the meaning and value derived from those consequences may be only relatively reliable. In this example, the child's level of experience and reflection is insufficient to enable an informed judgment of the possible consequences of eating chocolate, beyond the enjoyment expected. But it is this very factor that adults must acknowledge as essential to understanding how the child's judgment and behaviour is learnt and applied in various contexts. Indeed, this is how children do learn to discriminate among contexts where an activity is commonly deemed inappropriate. Before the child is capable of grasping the moral consequences that would render an act undesirable, the act is deliberately *made* undesirable by the punishment that would follow. Conversely, acts that are normally desirable in certain contexts are rewarded with affection. The capacity for self-control is thereby initially induced by the actual consequences that are produced in a variety of situations.

By immediately acting to satisfy the sudden desire for an Easter egg, the child is guided by its capacity to imagine the consequent enjoyment to be had. To that extent, the desired item is evaluated as 'good to eat', even though the expected value of the pleasure is as yet unlimited, owing to the child's ignorance of the social rules governing the acquisition of goods for private consumption. However, once an adult has warned the child of the severe punishment it will incur by taking goods without permission or payment, then the supermarket environment begins to exert a restraining influence upon the child's natural impulse for immediate gratification, while still continuing to provoke those grasping impulses by the seductive advertising and display of their goods. In this way, specific environmental conditions progressively work to define and refine the immediate evaluation of a desire, by exposing it to effects that variously heighten, diminish, or oppose its satisfaction by different degrees. Here, the

child begins with a crude, undifferentiated response based on universal expectations of enjoyment, as expressed in the unqualified, first -order judgment "good to eat". Those expectations soon prove to be decidedly optimistic, as the child is confronted by physical and social forces that prevent the free and full enjoyment of chocolate Easter eggs. At various stages of development, parental and social sanctions, seasonal availability and health costs are common examples of areas where the conditional consequences of consumption require a reassessment of the first-order value of absolute, unconditional enjoyment.

Only by actually acknowledging these limitations does the child learn to effectively control its immediate response to a stimulus. The act of witnessing the suffering of a sibling punished for eating an entire packet of chocolate biscuits may provide sufficient evidence for a child to resist the temptation to experience those consequences directly. Alternatively, it might attempt to circumvent those consequences by some deceitful strategy, such as blaming the sibling who has already proven to be untrustworthy. Likewise, the child in the supermarket may quickly learn to suppress the desire to take items from the shelves when supervised by parents, but may be willing to risk much greater punishment for stealing if circumstances make success seem likely. Thus, a child who succeeds in stealing an Easter egg in the company of a delinquent schoolmate will have little reason to regard such behaviour as wrong, until it becomes sufficiently aware of the actual negative consequences for others.[244]

Given that the social and material consequences of their actions are not properly understood by many young children, their level of self-control is typically due to a lack of knowledge and experience. It is only by learning the results of our actions in different contexts that enables an individual to judge the value of any first-order desire. Thus, there is no fallacy in requiring facts by which to determine the value of acting upon a desire in a given context. For it is only in the light of such information that desires, beliefs and values themselves *can* be formed and evaluated.

As reflective thought typically arises as a result of a disturbance in some prior activity, it is not a spontaneous or gratuitous process disconnected from experiences of the world. Even in Godfrey-Smith's example of boredom, that experience is still explicable in terms of some deficiency in the enjoyment which may be extracted from available means or ends-in-view encompassed in the immediate environment. Furthermore, even where those experiences are falsely imagined, they are nevertheless *experienced* as being produced by an actual object or event. Although it is not immediately aroused by any sensory disturbance perceived in the external world, the reflective thought that often occurs in dreaming is still provoked by problems experienced in the world. Similarly, even the most fantastic hallucinations and delusions are nevertheless perceived as emanating from an external source, from which the imagination distorts or conjures a false representation. It may be reassuring, or perhaps disconcerting to note that, for a time, even a philosopher as critically observant as Jean-Paul Sartre was disturbed by the belief that he was being "pursued by lobsters and crabs".[245] While the belief was based upon persistent hallucinations or "flashbacks" from his earlier experiments with the psychedelic drug mescaline, and probably triggered by the subsequent overuse of stimulants, those images were perceived as being caused by an external presence, not by a spontaneous internal projection. Even when suspecting the sensory evidence to be illusory, that suspicion will be warranted

or confirmed by some other external cause. Whether it be a from a drug, trauma, disease, dream state, or even a brain-in-a-vat, false beliefs are produced by an organic disturbance of external origin. By definition, false beliefs are not sufficiently justified by facts. But they are necessarily informed and induced by some actual experience of the world, however misconstrued.

(iii)   Dewey's Analysis of Reflective Thought

The analysis of fact/value relations and inductive reasoning in section (i) and the ontogeny of moral learning in section (ii) can now be used to better illustrate the dialectical process by which facts and values evolve in mutual dependence. The foregoing examples can thus be presented as they are produced in action and judged by inductive reasoning. This progression can be accurately expressed in the five essential phases identified in Dewey's analysis of reflective thinking.[246] As functional states of thinking, rather than stages, they need not always operate in strict sequence. But, as the following analysis will show, each function is a dialectically logical response aimed at resolving difficulties encountered in the progress towards some end. Engaged by these practical concerns, reflective thought involves:

(1) conceiving suggestions of possible solutions

(2) intellectualization of the problem

(3) conceiving an hypothesis to guide investigations

(4) inferential reasoning

(5) testing hypothesis by action (overt or imaginative)

Returning to the example of the ditch obstructing your path, Dewey noted that some suggested solution will immediately be envisaged e.g. to jump it (phase 1). However, conditions must be surveyed to confirm that they are conducive to success. These observations serve to focus intellectual attention upon the precise nature of the problem, so that it is unambiguously defined by specific elements of the situation (phase 2). Where, at first, the ditch was merely perceived as an obstacle, now it is specific features that are meaningfully relevant in relation to the plan to jump. As the width of the ditch and its slippery banks are recognised as facts counting against the success of that plan, it is discarded after further observations lead to the discovery of a log. That discovery then suggests the new possibility of building a bridge over the ditch (return to phase 1). With that end in mind, the size and weight of the log and its proximity to the ditch are the critical factors determining the probability of success. Having judged them to be adequate for the task, the new plan is no longer a mere suggestion, but now warrants serious consideration as a leading idea or hypothesis (phase 3). Guided by this definite idea, you are then able to infer the effects of various actions that may function as means of realizing it (phase 4). For example, you may compare the likely effects of pushing, pulling or rolling the log into position at various points and with various aids. Means are thereby selected and tested by results produced in the imagination, or in the actual situation itself (phase 5).

The process of reflective thinking continues to provoke and guide action until the immediate difficulty or doubt is resolved to the satisfaction of the enquirer. In this case, the problem posed by the ditch is literally past, by the act of crossing it. In addition, the selection and evaluation of available

means provides evidence for suggestions that may be applied to a range of different ends in different circumstances. For instance, logs that had been rejected as too short for crossing the ditch, may have been noted as potentially good firewood instead. Even while reflection is actively engaged in the search for means useful to a definite end-in-view, facts are continually being grasped and their value reassessed in relation to other possibilities anticipated. But those possibilities remain only sources for suggestions that lie dormant, at a prereflective level of consciousness, as yet untested by experience.

Until ideas have been tried and tested for their value as both means and ends, that value remains open to question. However, as common ideas and assumed values are developed and tested through the shared experiences of living, the individual soon comes to form certain definite convictions about the world. As long as these beliefs are able to survive exposure to factual necessities, they will not be falsified. As such, all beliefs remain hypotheses whose confirmation is never *absolutely* given, for the simple reason that past regularities are never a sufficient guarantee of future events. The most ancient of daily occurrences – the rising of the sun – provides scant evidence for its eventual extinction. However, the fact that we can reliably predict such changes by inferential reasoning does provide a *sufficient* reason to trust our judgments and act accordingly. The closer the results of our actions accord with our expectations, the more valuable those results in justifying those expectations. A good reason for action is therefore determined by the degree to which the consequences of that action can be reliably inferred, not by trusting past regularities which are liable to change, but by learning how phenomena behave under various influences and conditions. Where a wealth of experimental evidence gives a high degree of certainty to the consequences of an activity performed under certain conditions, then the actor has the foresight to anticipate and respond to possible obstacles, even to the extent of preventing their occurrence by altering the actual conditions that produce them.

For the same reason, desires, values and ideals that are fixed and unresponsive to changing conditions cannot be relied upon to produce results that are equally beneficial whenever and wherever followed. Well-being depends upon developing the practical intelligence to discern those facts which call for a desire to be restrained or satiated, or a value or ideal to be abandoned or upheld. Such questions are not simply concerned with identifying which facts are appropriate or potential objects or obstacles for a desire or value that is not itself in question. Unless desires and values are reflected upon in the light of prevailing conditions and probable consequences, they cannot be intelligently directed toward the individual's well-being. Even after facts and values have been pragmatically selected and assessed against each other, practical intelligence often involves controlling the degree of expression and motivation that a desire or value should exert in changing or maintaining the important facts of a situation. To that end, intelligence is also required in judging the degree of importance to ascribe to each relevant fact, insofar as it impedes or facilitates the achievement of a valued goal.

Moral judgment follows this dialectical process of reflection upon facts and values, until a hypothetical course of action is conceived and applied to the problem at hand. If, as Rawls maintains, reflective equilibrium is truly aimed at establishing an harmonious coherence between intuitive beliefs and values and those that are the result of ideally rational reflection, then facts cannot divested of their importance in informing that reflective choice. As action-guiding principles, moral concepts derive their significance from the results that they prescribe to effect. Whether it be aimed at utilitarian

consequences, moral character or virtue, or even the selection of principles specially designed to deter motivation by desired effects, it is the actions prescribed by a moral theory that give it value. But what are these actions aimed at, if not to maintain or transform the *facts* of a situation to serve some valued end? Conversely, how can the value of a given end be assessed, if not by observing how facts might help or hinder its completion? Thus, the choice of values is arrived at by reflection upon the relevant facts, which are in turn selected as relevant to some desired end.

Dialectical self-examination and practical testing of facts and values is driven by the need to rectify some imbalance between the way the world *is* and the way it *ought* to be, from the standpoint of the individual actor's informed preferences. Whether alone or allied with others, reflective activity is motivated by the subjective interests of the agent. But the expression and satisfaction of interests is frequently challenged by factual demands, which may be more or less intractable. Prominent among social facts are the normative values prescribed in various social contexts, values which are all too easily replicated by rewarding those who comply and punishing those who do not. Principles governing the exchange and distribution of private and public goods are facts whose normative acceptance is forcibly maintained by law, as such principles provide the foundation for the co-operative production of resources ensuring the well-being of individuals. As discussed earlier, implicit within these economic principles is the moral principle of reciprocal altruism.

The ancient belief that it is necessary to give in order to receive is a perfectly adaptive response to the persistent problem of scarcity. However, to the child presented with an abundant array of sweets on the supermarket shelves, neither scarcity nor private ownership would be grasped as factors that would effectively inhibit the desire to take anything. As thus far the child's understanding of reciprocity has been learnt and practised in the home, its meaning and value is specifically associated with that environment. While parents generally provide unconditional affection, guidance and protection, from an early age the child learns the value of controlling its behaviour in return for some reward. But if the child is not yet aware of the undesirable consequences of eating too much chocolate, or taking goods without paying, it is this lack of factual knowledge that prevents the development of the moral consciousness that would allow the child to gradually control its own behaviour in a variety of social situations. For the child lacking such knowledge, the process of reflection is substantially shorter (and in this case, sweeter) than a conscientious adult undergoes. Like the person confronting the ditch, the child's attention is aroused by a definite object. However, in this case the object is immediately apprehended as strongly desirable, so that its delectation and consumption constitute the unified end of the activity. As the child is ignorant of any undesirable consequences that might detract from the *prima facie* value of the end, the problem is only concerned with means i.e. devising and executing an efficient plan of action to realize the end. The movement from suggestion to distinct hypothesis is correspondingly reduced by that whole-hearted appraisal. As a result, the intellectualization of the problem, that would ideally include the assessment of the aforementioned consequences, is severely misguided. Premised upon an hypothesis that has not yet been informed by sufficient exposure to different social environments, inferences that ought to count against a course of action in a new situation will also be absent from reflection, leading to recommendations that are bound to produce nasty surprises.

There may be a few obstacles to negotiate – shopping trolleys, other people – but the plan of action is clear and direct. The child makes a bee-line for the largest Easter egg in sight and tears off the wrapping. But already an unforeseen difficulty presents itself, as the egg proves to be too large to bite. Trying to imagine a better approach, the child remembers seeing an older, more experienced sibling eating pieces broken up in a bowl. Without knowing precisely how the egg had been broken, the child can only infer, from its own frequent experience, that things can easily break apart when dropped, or hit against something. So the child throws the egg to the ground, where it breaks into dozens of pieces scattered along the aisle. Unfortunately, before the child can manage to collect them, they are crushed and sullied by a sudden rush of feet and wheels, some of which belong to the child's annoyed mother. Inevitably, it all ends in tears, as a severe reprimand is issued and sanctions loom for future misbehaviour.

In this case, unknown social facts prevent the child from making an informed judgment of the value of seeking immediate fulfilment of a desired end. But such disconcerting mistakes on the part of the child also produce greater unforeseen benefits for both parent and child. In noting the dangerous naïvety of the child's actions, the parent is forcibly reminded of the need to prepare the child for unfamiliar social environments that require self-restraint. Although too young to appreciate the economic or moral reasons for respecting private property, the child should at least be made aware of the *fact* that failure to do so is likely to bring undesirable consequences upon itself. In making this judgment of the *objective value* of such knowledge for the child, the parent is also guided by accepting the *fact* of being responsible for the child's well-being.

These evolving dialectical constraints upon moral reasoning provide substantial evidence against those 'analytical' theories which conceive values in 'intrinsic' terms, abstracted from the actual social context or agent-relative perspective from which all values arise and evolve. However, this critique will then furnish the basis for an *alternative* conception of intrinsic necessity which takes full account of the agent's own prudential interests.

(iv)   The Definist Fallacy

As these examples indicate, when the relations between facts and values are analysed within the concrete situations that give them a specific, practical function, the gap that appears in the purely abstract distinction is no longer relevant to the deliberative process. Given that values function to guide action towards ends which are conceived and planned in response to distinct factual conditions, they must be inferred in recognition of those facts which effectively function as premises restricting the practical cogency of bringing possible courses of action to a successful conclusion. As the bushwalking example indicates, *in practice* the theoretical gap between fact and value is often so constrained by the limited viability of accessible means/ends that it need hardly threaten the cogency of decision-making in most circumstances.[247] However, precisely because value judgments are hypotheses that are defined by their effects in a variety of contexts, they do not afford singularly determinate evaluations. For this simple reason, it is indeed a fallacy to identify goodness in terms of a singular characteristic, whether it be natural or non-natural. Thus, as Frankena notes,[248] it is the definist fallacy that Moore is really

concerned with, in targeting those philosophers who attempt to identify the non-natural quality of goodness with some natural characteristic or tendency. Although Moore uses the term "naturalistic fallacy" to refer to a typical version of the definist fallacy, his arguments are aimed at refuting the possibility of a definitive description of goodness in terms of any unique attribute, natural or metaphysical.[249] Nevertheless, while arguing against the identification of goodness with such a unique property, Moore was reluctant to "assert the more doubtful proposition that it is not identical with any whatever."[250] Amongst other admissions, Moore later noted that the error is not strictly a fallacy, as it involves no inference.[251] Rather, it rests upon a synthetic predication of value that is nevertheless held to be intrinsic or essential to the quality in question.

In this respect, Moore cites the proposition "Pleasure is good" to be a common error of this type, as it reduces goodness to a property that must be produced universally. Even if such definitions are only judgments about the effects of certain phenomena, Moore rightly dismisses the claim as a generalisation that is bound to be false, as the property in question is unlikely to produce the same good effects on every occasion.[252] Pleasurable activities need not always have good effects. Moore therefore agrees that the evaluation of good as a *means* is, after all, a hypothetical question, when its effects are considered under different conditions. But if "Pleasure is good" is interpreted as an assertion that pleasure is a good *end*, Moore objects that no contingent end can possess such intrinsic value. To be good in itself, pleasure would have to possess some property that is essential to the constitution of goodness as such.[253] However, as there are always alternative possibilities of intrinsic value, Moore states that "Accordingly, neither our judgements as to what actions we ought to perform, nor even our judgements as to the ends which they ought to produce, are pure judgements of intrinsic value."[254]

In other words, the evaluation of both means and ends can only be an evaluation of contingent possibilities of goodness, rather than an experience of its essential nature. Yet, such an observation only supports Mackie's contention that it is indeed an error to attribute such intrinsic value to ends that are in fact the expression of human interests. But Moore was committed to the Platonic view that goodness must partake of some ideal quality that necessarily transcends the mundane interests of mankind.[255] So instead of dispensing with the idea of intrinsic value and the impossibility of such "pure judgements", he was led to posit the existence of a special faculty of moral and aesthetic judgement, by which such values are intuitively known. For good measure, this error might be called the "intuitionistic fallacy", as Frankena suggests, using the term as loosely as Moore had.[256]

The belief that an end or activity can be 'good-in-itself' is often used to contrast with goods that have only instrumental value as means of obtaining this final end, whose value is somehow self-sustaining. Health and happiness are thus held to be valuable 'for their own sake', as if such ends were independent of our interest in them. However, if it is granted that values can only be produced and sustained in the thoughts and actions of the persons who put them into practice, no value can ever be good-in-itself. Precisely because goodness is a description of the *relation* of utility that an action or object bears towards a desired end, its meaning must be relative to that end. Yet, as Lemos rightly notes, "according to the traditional view, intrinsic value is a *nonrelational* concept" and also "according to this tradition, the intrinsic value of a thing is not dependent on its being the object of any psychological attitude."[257] In addition, Thomson notes that "a certain distinction on which proponents

of intrinsic value rely" is defined in terms of "nonderivative" and "derivative" goodness, which is also equated with a distinction between "intrinsic" and "extrinsic", as Zimmerman indicates.[258] According to this view, an extrinsic good derives its value from something else which may also be extrinsically good but these must be ultimately derived from some intrinsic i.e. nonderivative source of value. However, given that intrinsic value is understood as *nonrelational* and *independent* of any psychological attitude such as the prospect of enjoyment, then it is *impossible* that such enjoyment (or any other value satisfaction) could constitute a nonderivative source of value.

In the example above, the act of consuming the Easter egg is the end-in-view that gives value to the means employed to make it easy to consume. So the egg is broken only for the sake of the end, which is certainly thought to be good. But does the child really imagine eating Easter eggs to be an activity that is good-*in-itself*, having *no relation* to the desires of those who enjoy them? On the contrary, as the egg is desired for the sake of this enjoyment, its consumption is the means of delivering that pleasure. Perhaps, like health and happiness, here enjoyment might be *believed* to exhibit this mysterious intrinsic value, as the clear end of the activity. Yet, even if the end is *construed* as ultimate in defiance of the incongruity noted above, it is seldom valued without practical reference to other means and ends which are liable to diminish or heighten its realization. As these other means and ends also valued in terms of their *relation* to the 'ultimate' end, the same contradiction only reappears i.e. the claim that a value (intrinsic) can be the source of another value (extrinsic) yet have no psychological relation to that value.

In this case, the pleasure anticipated is specifically defined by an end-in-view. When the end is realised, the desire is satiated and the individual is given pause to redirect attention to other activities, all of which must be held as ends-in-view if they are to motivate action. It is important to note that the description 'end-in-view' does not preclude a long-range goal that may incorporate a series of ends as parts of the whole. In this sense, health, happiness, pleasure or general well-being are not evaluated independently of the means by which they are achievable. The pursuit of such global ends requires a considerable allocation and expenditure of resources that might be more profitably directed toward other activities. Therefore, the value of these end-states cannot reasonably be deemed final, or good-in-themselves, without considering the worth of the means invested in attaining them. Likewise, the efficacy of available means will limit the degree to which the end is approximated. As such, each prospective end is prioritised not only in relation to the others, but also in relation to the feasibility constraints involved in pursuing any particular end(s) on any particular occasion. Such constraints always ensure that there must be a finite set of possible means and ends which can find some satisfaction at any point in time. Therefore, there is no threat of any epistemic infinity or circularity in the pragmatic evaluation of ends-in-view. Furthermore, these mutual constraints among valued ends cannot be construed as conflicts among intrinsic values. As they arise by comparison of their value *in relation to each other*, they are *not* assessed independently of each other's satisfaction, which therefore disqualifies them as intrinsic values. Thus, in the case of a person who foregoes some prospect of pleasure because its attainment will hurt some loved one, neither of these conflicting values are held independently of the person's psychological attitude. On the contrary, it is only *from* that attitude of concern that the one values the other.

If final ends are to be distinguished from ends-in-view, there can be no further end to which they are directed as means. In arguing that such ends as honour, pleasure, intelligence and various virtues are valuable for their own sake, Aristotle admits that they are not absolutely final, as they are also chosen as means of securing well-being.[259] It is only the latter that has no other purpose and is thus self-sufficient. Similarly, Hume argues thus: "Ask a man why he uses exercise; he will answer *because he desires to keep his health*. If you then enquire *why he desires health*, he will readily reply *because sickness is painful*. If you push your enquiries further and desire a reason *why he hates pain*, it is impossible he can ever give any."[260] Hume's account of ultimate ends anticipates the basis of Bentham's utilitarianism, where avoidance of pain and the pursuit of pleasure are the ultimate ends of human activity. For Hume, pleasure is desirable in itself "because of its immediate accord or agreement with human sentiment and affection."[261] By contrast, in defending happiness as the ultimate end, Mill is careful to include the following, important proviso: "No reason can be given why the general happiness is desirable, except that each person, *so far as he believes it to be attainable*, desires his own happiness." (Italics mine).[262]

Mill acknowledges the desirability of happiness to be conditioned by the probability of attaining it. In so doing, he risks undermining the very foundation of his argument for happiness as the ultimate end. Although the context of the statement indicates that Mill wants to emphasise that general happiness must reflect the private judgements of individuals, he nevertheless grants that happiness cannot carry the unconditional, universal assent that it requires as the ultimate end of desire. As the general happiness is attainable only by satisfying the desires of individuals, it must be built upon a variety of judgments concerning the happiness of each individual life. For this reason, neither health, pleasure, intelligence, nor any other end can be claimed to admit no further evaluation. Indeed, even if it is well-being rather than happiness that is universally sought, it is not pursued as a final *end* of activity. If that were the case, its attainment would then render all activity pointless, paradoxically including the very activities that promote well-being. Therefore, rather than comprising a final, sufficient end for the fulfilment of all other ends-in-view, well-being must function as a condition for some further activity. Of course, even the most contented potentate can remain so only by undertaking further actions in the interests of his future well-being or happiness. Indeed, if well-being is conceived as an end-state of life activities, its pursuit would have no survival value. But once it is recognised that states of well-being serve as necessary conditions for the maintenance of such purposive activities, their value becomes clear and can here be described as intrinsic by the more common, alternative meaning synonymous with 'essential' or 'indispensable'. Thus, in addition to the natural selection of autonomy, well-being is a functional condition that enables the organism to survive and reproduce. Furthermore, the co-evolution of these two functions produce a mutually enhancing synergy, so that each is vital for the development of the other. Autonomy allows the organism to explore a range of possible means of maintaining well-being. In turn, enhanced well-being heightens the organism's capacity for exercising its autonomy. As the next chapter will argue, in human functioning a certain level of reflective autonomy and well-being are essential conditions for the development of moral agency.

---

[184] Moore, op. cit., p.46ff

[185] Hare, *The Language of Morals*, pp.118-126

[186] W. V. Quine, "Two Dogmas of Empiricism," in *From a Logical Point of View: Nine Logico-Philosophical Essays* (Cambridge, Mass., Harvard University Press, 1961). pp.20-46

[187] Hilary Putnam, *The Collapse of the Fact/Value Dichotomy and Other Essays* (Cambridge Mass. & London, England, Harvard University Press, 2002), p.14

[188] Ibid., p.15

[189] Quine, "Two Dogmas of Empiricism," p.31

[190] David Hume, *Enquiries Concerning Human Understanding*, ed. P. H. Nidditch & L. A. Selby-Bigge (Oxford, Clarendon Press, 1975), IV. I

[191] John Dewey, "The Evolutionary Method as Applied to Morality," in *The Middle Works,* vol. 2, p.38

[192] Ibid., p.32.

[193] Ibid.

[194] Ibid.

[195] Even so, Locke himself nevertheless regards the nominal essence of moral and practical ideas as constituting their *real* essence: see Stephen Darwall, *The British Moralists and the Internal 'Ought': 1640-1740* (Cambridge, Cambridge University Press, 1995), p.44

[196] Ibid., pp.44-45

[197] Mark Platts, *Ways of Meaning* (London, Routledge & Kegan Paul, 1979) p.257, cited in I.L. Humberstone, "Direction of Fit," *Mind* 101 (1992), p.59

[198] Putnam, op.cit., pp.9-11; See also Gerhard Shurz, *The Is-Ought Problem*: *An Investigation in Philosophical Logic* (Dordrecht, Netherlands; Kluwer Academic Publishers, 1997), p.25

[199] Humberstone, op. cit., p.70

[200] Hare, *Language of Morals*, p.81

[201] Nicholas Rescher, *A System of Pragmatic Idealism: Volume II: The Validity of Values*: *A Normative Theory of Evaluative Rationality* (Princeton, Princeton University Press, 1993), pp.81-84

[202] Ibid., p.84

[203] A. C. MacIntyre, "Hume on 'is' and 'ought'" in W. D. Hudson (ed.) *The Is-Ought Question: A Collection of Papers on the Central Problem in Moral Philosophy* (London, Macmillan, 1969), pp.36-37. It should be noted that this interpretation does not depend upon also subscribing to MacIntyre's more dubious contentions regarding either Hume's understanding of 'deduction', or that it is only 'vulgar' systems of morality which Hume criticises on these grounds. See R. F. Atkinson, "Hume on 'is' and 'ought': A Reply to Mr MacIntyre," in Hudson, op. cit., pp.53-58

[204] Ibid., pp.40-42

[205] Schurz, op. cit., p.31; see also MacIntyre, "Hume on 'is' and 'ought'", p.79

[206] John Dewey, *How We Think : A Restatement of the Relation of Reflective Thinking to the Educative Process*, rev. ed. (Boston ; New York, Heath. 1933). p.107

[207] Ibid., p.105

[208] Cf. Heidegger's phenomenological descriptions of objects perceived as "ready-to-hand" vs. "present-at-hand" in *Being and Time* trans. J. Macquarrie & E. Robinson (London, SCM Press, 1962) pp.96-7, 99 & 103; & esp. Sartre's example in *Notebooks for an Ethics*, trans. D. Pellauer (Chicago,

University of Chicago Press, 1992), pp.97-9, where it is the *absence* of a plank which reveals its necessity as a means for the end of crossing a gully

[209] John Dewey, "Value , Objective Reference, and Criticism," (1925) in *The Essential Dewey* vol. 2, ed. L. Hickman & T. M. Alexander, p.291

[210] See John Dewey, "The Logic of Judgments of Practice," in *The Essential Dewey* vol. 2, pp.245-246

[211] John Dewey, *Theory of Valuation*, (Chicago, The University Of Chicago Press, 1939), p.36

[212] Jean Piaget, "Problems of Equilibration," in *The Essential Piaget*, p.839

[213] Jean Piaget, "Equilibration Processes in the Psychobiological Development of the Child," in *The Essential Piaget*, p.836

[214] See Robert J. Richards, *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior* (Chicago & London, The University of Chicago Press, 1987), p.464, note 46 on Piaget's acknowledgement of Baldwin's influence; and p.467-475 on the phylogenetic effects of Baldwin's ontogenetic principles

[215] For seminal work in domain-specific paradigm see Elliot Turiel, *The Development of Social Knowledge: Morality and Convention* (Cambridge, Cambridge University Press, 1983)

[216] For summation of stages in both theories see Lawrence Kohlberg, *The Philosophy of Moral Development: Moral Stages and the Idea of Justice* (San Francisco; London, Harper & Row, 1981), pp.17-20; Jean Piaget, *The Moral Judgment of the Child* (Routledge & Kegan Paul, 1932), pp.314-316 (or extract in *The Essential Piaget*, p.187-182)

[217] Kohlberg, op.cit., p.16

[218] Ibid., pp.23-26. See Lawrence J. Walker et al., "Reasoning about Morality and Real-Life Moral Problems," in M. Killen & D. Hart (eds.) *Morality in Everyday Life: Developmental Perspectives* (Cambridge; New York, Cambridge University Press, 1995), pp.374, 389 &403.

[219] M. Laupa, E. Turiel & P. A. Cowan, "Obedience to Authority in Children and Adults," in *Morality in Everyday Life*, pp.136-137

[220] Judith G. Smetana, "Parenting and the Development of Social Knowledge Reconceptualized," in J. E. Grusec & L. Kuczynski (eds.) *Parenting and Children's Internalization of Values: A Handbook of Contemporary Theory* (New York, John Wiley & Sons, 1997), pp.167-168

[221] Ibid., pp.172 & 177

[222] Ibid., p.164

[223] Smetana, "Context, Conflict and Constraint in Adolescent-Parent Authority Relationships," in *Morality in Everyday Life*, pp.230-231; Cf. Ch. 1 of *The Moral Judgment of the Child* for Piaget's classic account of children's evolving conceptions of the authority and function of rules in games of marbles and hide-and-seek

[224] Smetana, ibid., p.231

[225] Piaget, *The Moral Judgment of the Child*, p.316

[226] Orlando Lourenço, "The Aretaic Domain and Its Relation to the Deontic Domain in Moral Reasoning," in M. Laupa (ed.) *Rights and Wrongs: How Children and Young Adults Evaluate the World* (San Francisco, CA; Jossey-Bass, 2000), p.49

[227] L. Kuczynski, S. Marshall & K. Schell, "Value Socialization in Bidirectional Context," in *Parenting and Children's Internalization of Values*, p.38

[228] D. Hart & M. Killen, "Perspectives on Morality in Everyday Life," in *Morality in Everyday Life*, p.9

[229] For supporting experimental evidence see Charles C. Helwig, "Social Context in Social Cognition: Psychological Harm and Civil Liberties," in *Morality in Everyday Life*, pp.177-192; see also Laupa, Turiel & Cowan, op.cit., pp.155-158

[230] Howard E. Gruber & J. Jacques Vonèche, "Looking Back at the Essential," in *The Essential Piaget*, p.866

[231] Ibid.

[232] Richards, op.cit., pp.464-469

[233] Kuczinski et al., op.cit., p.33

[234] Ibid., p.30; cf. Piaget note 33 above.

[235] Hart & Killen, op.cit., p.7

[236] Ibid., xiii

[237] Moreover, anthropological evidence from a range of different cultures supports Vygotsky's contention that such dramatisation and role playing facilitates the development of autonomy in enabling children to experiment with imaginative possibilities unconstrained by current circumstances: see Barbara Rogoff, *The Cultural Nature of Human Development* (Oxford & New York, Oxford University Press, 2003), pp.298-299

[238] See selections of Piaget's experiments from "The Child's Conception of Movement and Speed," and esp. "The Child's Conception of Time" in *The Essential Piaget*, pp. 547-563 showing clear stages in the child's intuitions in differentiating between space and time e.g. at one stage confusing time and velocity by the following reasoning: "(1) If you go more quickly you necessarily cover more space (i.e., velocity is proportional to distance); (2) if you cover more space, you need more time to do so (hence distance is proportional to time); and (3) if you go more quickly you need more time because you cover more space." p.556.

[239] From a pragmatic perspective this explains the economic and moral function of myths and fairytales as means of controlling children's innate selfishness until they can be reasoned with. Cf. Jacqueline J. Goodnow, "Parenting and the Transmission and Internalization of Values: From Social-Cultural Perspectives to Within-Family Analyses," in *Parenting and Children's Internalization of Values*, p.341 on clear evidence that "parents encourage some forms of storytelling, setting others aside as uninteresting or as too "fanciful," sometimes to the point of labelling as the telling of "lies" the stories that schoolteachers regard as "imaginative".

[240] See Wendy S. Grolnik, Edward L. Deci & Richard M. Ryan, "Internalization within the Family: The Self-Determination Theory Perspective," in *Parenting and Children's Internalization of Values*, pp.143-146

[241] It must be admitted that Dewey is uncharacteristically dogmatic in insisting that the instrumental role of enquiry should always be aimed at resolving some indeterminacy. As Godfrey-Smith observes in *Complexity and the Function of Mind in Nature* (Cambridge, Cambridge University Press, 1996), p.112, such indeterminacy may sometimes be precisely what the agent seeks, as suggested by the

experience of boredom in response to predictably determinate features of the environment. Though Dewey may well reply somewhat paradoxically that the 'arousal' of boredom then constitutes a new problem to be given a determinate solution, that solution can scarcely be found in features of the immediate environment which are perceived as fixed in their potentialities.

[242] Dewey, *Theory of Valuation*, p.21

[243] Hare, *Freedom and Reason*, (Oxford, Oxford University Press, 1963), pp.18-20

[244] See Lawrence J. Walker et al., "Reasoning about Morality and Real-Life Moral Problems," in *Morality in Everyday Life*, p.381-382 for evidence that individuals of various ages still commonly invoke prudential considerations "within a moral framework" as grounds for or against possible actions e.g. "losing one's job or a major client, getting arrested for shoplifting or expelled from school" vs. "anticipated benefits (e.g., one's own pleasure, eventual rewards)."

[245] See Annie Coen-Solal, *Sartre: A Life*, pp.104 & 382. Note: This incident may well have been the inspiration for Monty Python's "Spiny Norman" – see Python Productions Ltd., *Monty Python's Big Red Book* (London, Eyre Methuen, 1972)

[246] Dewey, *How We Think*, p.107

[247] Cf. Singer's discussion re "The Triviality of the Debate over "Is-Ought" and the Definition of "Moral" in American Philosophical Quarterly, 1973, vol.10, pp.51-56, which focuses upon the limits to which definitional disagreements over logical accounts of morality should be allowed to set the agenda in moral philosophy. Singer's point would be well directed at the sterility of the efforts of those logical positivists who are largely responsible for invoking Hume's 'Law' to justify their dismissal of any attempt to defend moral reasoning on pragmatic, normative grounds. As Dewey notes, "In terms of morality, "the problem is not one of fact but of value. It is a logical problem. If we suppose such necessary and universal beliefs as go by the name of "intuition" to exist, does such existence settle anything regarding the validity of what is believed, either in general or in part?" – "The Evolutionary Method as Applied to Morality," p.25

[248] W. K. Frankena, "The Naturalistic Fallacy," in *Theories of Ethics*, ed. P. Foot (Oxford, Oxford University Press, 1967), p.57

[249] Moore, op. cit., p.13

[250] Ibid., p.15

[251] Ibid., pp. 20-21

[252] Ibid., p.22 & pp.74-5

[253] Ibid., p.22

[254] Ibid., p.77

[255] Ibid., xxxiii

[256] Frankena, op. cit., p.61

[257] Noah M. Lemos, *Intrinsic Value: Concept and Warrant* (New York, Cambridge University Press, 1994), pp.3 & 5

[258] Judith Jarvis Thomson, "On Some Ways in Which a Thing Can Be Good," in E. F. Paul et al (eds.) *The Good Life and the Human Good* (Cambridge, Cambridge University Press, 1992) pp.99ff., cited in Michael J. Zimmerman, *The Nature of Intrinsic Value* (Lanham and Oxford, Rowman & Littlefield,

2001), pp.22-23

[259] Aristotle, *The Nicomachean Ethics*, I. vii. 4-5

[260] Hume, *Enquiries Concerning the Principles of Morals* Appendix 1, v

[261] Ibid.

[262] John Stuart Mill, *Utilitarianism*, in *Utilitarianism; On Liberty; Considerations on Representative Government; Remarks on Bentham's Philosophy*, ed. G. Williams (London, Everyman, 1993), pp.32-33.

CHAPTER FIVE

THE EVOLUTIONARY CONDITIONS OF AGENCY

As argued in the previous chapter, as a reflective process, moral reasoning is largely guided by inductive principles and standards of justification. These synthetic norms do not claim to establish any finality or certainty in the judgements they support, but only to advance the agent's knowledge of the probable consequences of pursuing different means and ends in various contexts. Thus, it was shown that inductive and instrumental reasoning from facts to values need not violate any pragmatically rational norms. Likewise, it was argued that such judgements need not be premised upon any preconceived theory of final ends or 'intrinsic' values independent of agents' interests. Such arguments might seem to imply that no moral claim can ever be guaranteed by deductive certainty. Yet, as will be shown in the following section, the conditions by which human agency is *necessarily* realized can constitute the essential (though provisional) foundations required for a deductive, categorical defence of universal rights to the fulfilment of those conditions.

(i)   Gewirth and the Conditions of Agency

Alan Gewirth has presented the most sustained, coherent argument to demonstrate that both freedom and well-being are imperatives in the teleological functioning of human life. Therefore, he concludes, they must be valued as necessary goods in enabling purposive agents to pursue their chosen goals.[263] One of the strengths of Gewirth's method is that, like Dewey's, its reasoning proceeds from the subjective, first-person perspective of the agent, rather than some external 'ideal observer' who has no direct insight into the motives and preferences of individual actors. Instead, as long as we conceive of ourselves as prospective purposive agents (PPA), we can readily identify with the reasoning that follows. The argument involves several steps in the PPA's judgement, essentially as follows:[264]

(1) I do *X* for purpose *E*.

(2) *E* is good.

(3) My freedom and well-being are necessary goods (for my agency).

(4) I have rights to freedom and well-being.

(5) All prospective purposive agents have rights to freedom and well-being.

The first two steps are uncontroversial reports of the agent's (1) recognition of acting for a purpose and (2) regarding that purpose as good. The scope for misjudging that or any other purpose as good has no bearing on the psychological fact that agents must ordinarily regard their chosen goals as worth pursuing. Likewise, the phenomenon of *akrasia* is irrelevant to the agent's present belief that *E* is good. Step (3) simply affirms freedom and well-being as the necessary conditions for the possibility of (1) and (2). The extent to which either is necessary for full agency may be difficult to establish. The claim here is limited to acknowledging their function as the *sine qua non* of human agency. In that respect, it is a factual claim. Yet, as it also affirms freedom and well-being as necessary *goods*, it a judgement of their essential value for human agency. As such, it follows that an agent must be able to claim these goods in order to function as a PPA. Thus I have rights to freedom and well-being (4). The final step (5) simply extends the recognition of those first-person rights to all other prospective, purposive agents, on the grounds of generic consistency. It is this "Principle of Generic Consistency" by which is/ought, fact/value categories are inherently connected to the principle of necessity.[265] Firstly, from the agent's own experience, freedom and well-being are inferred as necessary means for the agent's purposes to be attainable, whatever they may be. By the same reasoning, the agent is committed to drawing the same conclusion about other persons with the same generic features of agency. It is only in virtue of being a PPA that the agent can claim (4) "I have rights to freedom and well-being". From the logical principle of non-contradiction, the agent cannot then rationally deny that other PPAs have those same rights.

From the earlier reconstruction of moral value, the evolutionary function of common evaluative terms was shown to be primarily prescriptive in proffering advice, but also descriptive in its evaluation of objective phenomena. In giving directions to a hungry stranger, the "right" way to a restaurant for a "good" meal was cited as evidence for equating such terms with evaluations of utility that are nevertheless objective in the light of the goal expressed by the stranger. What the goal *is* necessarily delineates which actions *ought* to be taken to realize it. The act of eating derives its objective *value* from the *fact* that the person is hungry. To be sure, as noted earlier, it is not possible to *deduce* any value from the fact that a person has a desire. Moore is right to insist that it remains an open question how one might evaluate any desire in the absence of any fixed standard of goodness. Likewise, Frankfurt's "unwilling addict" can be cited to illustrate that the mere possession of a desire cannot require that its satisfaction be valued.[266] Likewise, the fact that I have a biological need to eat does not entail that I should value eating at all. I may well regard it as an unpleasant chore, even to the point where I prefer to starve to death rather than satisfy my hunger.

Given that ascriptions of value are subjectively determined by the preferences of individuals, it would seem that no objective knowledge can necessarily determine any definitive evaluation of that knowledge. This recognition of the autonomy of personal judgment was the basis of the earlier

argument *against* the possibility of a categorical imperative. But, conversely, in showing autonomy to be a *necessary* condition for the possibility of making any *subjective* value judgment, its *objective* value for human agency is thereby established. This kind of argument is typically described as "transcendental", in recognition of its central role in Kant's attempt to "deduce" the necessity of certain *a priori* categories of thought that must be present for our experience of the world to be possible.[267] However, as the above argument indicates, the essential value of autonomy need not depend upon any faith in the workings of metaphysics. Kant himself saw that our experience of autonomy in practice bears witness to its vital function, however much we may doubt its priority in the "spontaneity" of our thought processes.[268]

By virtue of the will's dependence on autonomy, it is objectively necessary to act from obligation.[269] Autonomy is therefore the necessary condition for having a good will. However, unless it can also be shown to be a *sufficient* condition for generating acts that can only be good, the autonomous will cannot require categorical obedience. If we are deprived of the ability to assess and respond to the consequences of our actions, we thereby deprive ourselves of the autonomy on which our moral judgments depend. As Gewirth argues, while a certain degree of freedom is categorically required as a precondition for any purposive activity, a minimal level of well-being is also necessary for those purposes to be realised.[270] Without such basic material and psychological conditions as are necessary for maintaining the ability to function as prospective, purposive agents, autonomy is denied the capacity to even conceive of such acts as a good will might allow. As such, 'pure' reason lacks the practical demonstration that Kant claims to 'deduce' from the concept of a being with a will i.e. that such a being must necessarily act under the presumption of freedom.

By contrast, Gewirth reaches the same conclusion by a conceptual analysis of the generic features of purposive action that any agent must rationally accept as necessary conditions. In reasoning from the standpoint of the agent, Gewirth's dialectically necessary method conceives the agent's judgments or "statements" as "necessary ones in that they reflect what is conceptually necessary to being an agent who voluntarily or freely acts for purposes he wants to attain."[271] Thus, as Gewirth himself notes, this conceptual analysis also has a "demonstrative" aspect in that such necessities in the "nature of action reflect the necessities of this existing world, including the limits set by its own structure and potentialities."[272] Also, the conceptual analysis itself proceeds from the "empirical premiss" of "considering actions as they are viewed and referred to by the agent himself."[273] For this reason, Gewirth's conclusions have firm foundations in the practical nature of human reason and freedom, whose functioning is dependent upon the existence of certain material goods. These essential goods provide a specific content to defining moral imperatives in terms of self-preservation. Nevertheless, insofar as such goods are only aimed at satisfying human interests, Kant would reject them as candidates for moral value, as that value remains determined by laws of nature.[274] Yet, in arguing that genuine autonomy can only be exercised in opposition to natural inclinations, Kant thereby implies that it must in fact be permitted by the laws of nature. If moral autonomy consists in "the fitness of the maxim of every good will to make itself a universal law",[275] then that fitness cannot be determined by reason alone. It must also be accommodated by human propensities, on the basis of Kant's own principle that 'ought' implies 'can'. However, while admitting that anthropological data are

required to guide the application of moral principles, Kant still insists that the principles themselves must be independent of any contingent human desires or interests.[276]

Kant failed to recognise that not all human desires or interests need be contingent. As Gewirth's analysis reveals, autonomy need not be conceived as an ideal opposed to human nature, but as one required by practical reason as instrumental in maintaining and developing that nature. Beyleveld neatly distinguishes Kant's conception of reason requiring a supreme moral principle to be "*independent* of all contingent willings", as opposed to Gewirth's requiring only that it "sets requirements *irrespective* of what might be contingently willed."[277] Where Kant's foundationalism only succeeds in severing all links to the interests or desires that evolved to motivate action, Gewirth's focus on the conditions of agency reveals the essential unity of reason and value. A prospective purposive agent has an undeniably practical reason to value the conditions of its agency. But is it necessarily true that the agent must arrive at such a positive evaluation? Would it not be possible to regard one's freedom and well-being with indifference or outright contempt, even while acting on its behalf? Yet again, the wanton, the willing and the unwilling addict might be cited to refute any categorical or logical requirement in the evaluation of a particular desire. But it would be a serious mistake to interpret the expression of this valuational freedom as evidence *against* its necessity. The fact that one *can choose* to despise or neglect one's freedom and well-being is only possible if those elements are implicitly acknowledged as "necessary goods" for one's purposes. Thus, even the decision to renounce one's status as a prospective purposive agent cannot logically deny the objective value of the freedom and well-being that is necessary to consistently avoid undertaking any further reflective, purposive action in future. Given that such intentional avoidance of purposive activity can scarcely be attempted without any deliberate action to thwart any further interest in one's welfare, those efforts are sufficient to attest to the actual value of the freedom and well-being they are designed to resist.

The person who no longer desires to desire does not pose any challenge to Gewirth's contention that freedom and well-being are vital for prospective purposive agents. It is not necessary for the agent to always act on the basis of a conative appraisal of these conditions as necessary goods. After all, we do not ordinarily reflect upon the generic features involved in our purposive actions. However, Regan stresses this point to argue that Gewirth's conception of agency is misconstrued in attributing these generic features to the agent's particular purposes.[278] In effect, this objection contests the validity of inferring that (3) My freedom and well-being are necessary goods (for my agency) simply on the grounds of affirming that (1) I do *X* for purpose *E* and (2) *E* is good. The belief in the value of a particular activity need only entail that one must will the conditions necessary for *its* articulation, rather than the generic conditions that make *all* purposes practicable. For instance, Regan argues that one who only desires to excel at playing the piano need not be committed to valuing the freedom required to perform any other activities.[279] Even if such a person is likely to value the musical skills that can be practised on alternative instruments, or must admit to valuing creativity generally, the approval of this range of activities will not provide sufficient grounds for appreciating the freedom to play video games, as long as that activity is regarded as "a seductive and stultifying waste of time".[280]

The goods that an agent must rationally admit as necessary in the implementation and maintenance of any particular activity need only be premissed upon a conception of the good that is

confined to the agent's prevailing concerns. What of the agent who is committed to an ideal or practice so highly esteemed as to render all unrelated activities meaningless? What of the pianist who literally lives only to play, so that if he were to lose the use of his hands, neither freedom nor well-being could have any further value? Further to this point, Regan cites Bernard Williams' distinction between projects that people are committed to unconditionally, for the sake of which they desire to live, and those they pursue only so long as life permits.[281] In that case, the necessity of freedom and well-being would seem to be directly dependent upon the relative importance of the agent's projects which, in turn, will be conditioned by the opportunities afforded to implement them.

This conditional relativity might be seen as evidence supporting Kant's dismissal of any rational necessity based on the contingency of an agent's personal desires. However, the psychological fact that agents invest their particular, individual projects with varying degrees of concern is perfectly compatible with Gewirth's claim that an agent must be rationally committed to valuing the generic conditions needed to undertake such projects. To be sure, an agent need not explicitly infer that freedom and well-being are generically necessary for embarking upon any conceivable project. But these criticisms are misdirected at Gewirth's definition of the agent as one who is concerned to maintain the capacity to act upon further choices in future. As such *prospective* agency consists in both the functional capacity to choose alternative actions and the ability to implement them in future, it sets certain preconditions for the prospects of satisfying *any* contingent desires. Therefore, even though Regan is right to point out that a purposive agent need not reflect upon the generic grounds of its agency in order to pursue a particular end, if such an agent fails to become aware of its freedom to choose other ends, then it lacks the degree of inductive reasoning and self-awareness that is a prerequisite for practical and moral thinking in human societies.

This restricted agency is comparable to that exhibited by other primate species lacking the sheer cognitive capacity to infer general ideas from particular actions. The chimpanzee who fortuitously discovers that twigs are good for extracting termites from a hollow log, may learn to apply that knowledge to a different end – for example, in drawing honey from a beehive. But unlike the mature human agent, the chimpanzee lacks the imaginative ability to conceptualize that particular skill as one that is derived from the wider, general capacity to choose alternative possibilities. Indeed, by granting that human agents must, at least, reflect upon their contingent desires from some conception of the good, Regan effectively affirms this conceptual insight to be a necessary condition for some wider agency. But where Regan and others are content to assume that the question of justification begins and ends with no firm rational basis to these beliefs, Gewirth's analysis of the process of dialectical reasoning reveals the logic governing their coherence.

The dedicated pianist who values his freedom and well-being only to the extent that it affords him the scope to develop and maintain that skill, must still act upon the assumption that there should be sufficient conditions to do so in the foreseeable future. Therefore, even the process of identifying a singular source of value forces the agent to look beyond the immediate conditions to those that are absolutely essential to preserving that source. The pianist cannot rationally account for his freedom to play as if such freedom were somehow determined only by his interest in playing. If that were the case, he would not be aware of his freedom *not* to play, or to take up another instrument or activity. From the

very beginning, without an implicit awareness of his freedom as a generic, all-purpose function, he would not have been able to plan and execute all the other actions that serve as means for rendering the occasion to practice efficiently. This alone would entail the need to secure sufficient funds to cover the cost of lessons and materials etc. Even if the supply of such necessities may be fully met by others, even the most pampered, self-absorbed artist will appreciate having the ability to choose *this* particular keyboard, or to protect his precious hands with *that* brand of kid gloves.

All of the experiences involved in planning and continuing a chosen course of action inevitably testify to a conception of oneself as an autonomous agent. The same experiences also demand a measure of self-concern that presupposes one's recognition of the basic material goods that are needed to remain a prospective, purposive agent. However, in valuing a purpose the PPA is not required to be immediately cognisant of the epistemic principle that supports it. When the grounds for any hypothesis have reliably presented themselves to the enquiring mind, having dispelled all doubts about their causal sufficiency, they need no longer be called to mind before acting upon them. The passion and technical skill of the pianist presents an apt illustration of the process whereby conscientious attention to theory is gradually neglected, as practice is perfected to the point where theory can be forgotten. Indeed, an over-reliance on theory often proves to be an impediment to progress in most areas of life, for economic reasons alone.

Once learnt and perfected, most regular activities are performed "like riding a bicycle". Deliberate attention or analysis given to the mechanics of the task only disrupts an operation that would otherwise run very efficiently. In the same manner, the experienced, rational agent has little incentive to recall the elementary preconditions for his agency before putting it to use. Nevertheless, it might still be noted that many people appear to exercise full agency without ever consciously acknowledging freedom or well-being to be their necessary conditions. But this observation would ultimately have to be attested by the people concerned, which is precisely what Gewirth's dialectical approach aims to establish i.e. that rational agents must admit these conditions to be necessary goods, even if they have not previously had occasion to clearly identify them as such.

In spite of these arguments in support of the dialectical, agent-centred analysis of moral thinking, the criticisms advanced by Regan do undermine Gewirth's claim to have deduced the *generic* necessity of moral rights from the strictly logical necessity of adhering to the principle of non-contradiction. By showing that an agent's conception of their own freedom and well-being is crucially dependent upon the value assigned to their personal projects and interests, then consistency demands only that the agent affirm the value of such freedom and well-being as suffices to meet those ends. Thus, from the particular evaluation that "*E* is good", all that can be legitimately deduced is limited to acknowledging the value of whatever means are necessary for this activity. Gewirth's argument for the Principle of Generic Consistency, based as it is on a purely formal analysis of the logical structure of thought, succeeds in deriving unconditional, universal rights only by ignoring the specific relativity of an agent's conception of the good.

One of the virtues of Gewirth's approach is that it is able to reveal the generic conditions of agency without having to prejudge whether any particular desired end is achievable or coheres with the agent's other beliefs and values. In any isolated judgment that "*E* is good", both the content of *E* and

the evaluation of its goodness are irrelevant to Gewirth's aim of demonstrating that the pursuit of any valued end necessarily commits one to valuing the means to achieve it. But while freedom and well-being are certainly indispensable means for whatever purposes an agent *might* have, that is not a conclusion that follows directly from the purposes that an individual *actually* entertains at any given time. Even if it is possible that a prospective, purposive agent might imagine, like Dr. Pangloss in Voltaire's *Candide*, that everything is good in the "best of all possible worlds",[282] the conditions of life impose practical limits upon the pursuit of some or even all of the activities that an agent might reasonably conceive to be good.

Inevitably, an agent's claim to freedom and well-being is not a claim to require an abstract, unlimited power to do whatever is willed to be a desirable end. As Dewey's analysis showed, the desirability of ends and means is judged by the agent's rational assessment of the degree of difficulty they present in the prevailing conditions. In this respect, although Gewirth's analysis of moral reasoning attempts to preserve the agent-relativity of value judgments that was neglected in Moore's abstract conception of intrinsic value, he nevertheless commits the same error as Moore by ignoring the hypothetical nature of such judgments. While all particular evaluations of goodness can be represented in the propositional form "*E* is good", this semantic simplification takes no account of the degree of belief invested in the worthiness of that particular *E*. Far from being irrelevant, these nuances are crucial in justifying the inference that Gewirth claims to be a matter of purely formal entailment. Rather than being categorically required for all conceivable purposes, freedom and well-being will be more or less valuable to an agent, depending on the level of commitment to the end being assessed which, in turn, will be influenced by the efficacy of means available. The agent's physical and mental capacities will also be an important factor in determining the level of freedom and well-being that must be deemed essential for pursuing any particular activity.

All such factors play a significant role in the formation of beliefs about value, indicating the relative indeterminacy which necessarily qualifies any expression of certainty that "*E* is good". As a proposition that is derived from a particular agent's set of experiences, any singular value judgment will have a conditional status, in that it is continually subject to reassessment in the light of the agent's further experiences. It is this prospective feature of agency that Gewirth duly seeks to preserve as a right for the agent to remain capable of free action in future. But such a right to freedom and well-being is therefore contingent upon the level of the agent's concern for continuing free action, a concern which can never be adequately justified, let alone deduced on the basis of any *current* concern for some particular activity. Therefore, strictly speaking, prospective, purposive agents may individually claim a categorical right to freedom and well-being only on the *hypothetical* condition that they actually value their agency.[283] As such, no PPA can be justified in claiming these goods to be *unconditionally* required for *any* purpose which *might* be valued in future. Furthermore, as a condition for continuing agency, this prospective interest constitutes a classically *inductive* argument in its reasoning toward future possibilities which must remain uncertain. This inductive uncertainty in the continuing value of specific goals thus restricts the scope of categorical rights which may be derived from the Principle of Generic Consistency on purely deductive grounds. Goods which may be deductively inferred as essential for the

pursuit of purposes which have only *prospective* value would only entail a right to the agent's effective freedom of choice and such minimal goods as are needed to sustain that efficacy.

Most of the criticisms of Gewirth's argument question the legitimacy of the PPA claiming a moral right to their freedom and well-being on the basis that such goods are at best *prudentially* required for the agent's purposes. Essentially, there are two common objections which contend that: (1) rights claims cannot be deduced from prudential requirements and that (2) moral rights cannot be deduced from prudential rights. Thus, even if (1) is logically defensible (2) need not be, and of course if (1) cannot be established, then neither can (2). Nevertheless, as Boylan suggests, almost all of Gewirth's critics have not adequately grasped the *dialectical* structure of the argument which takes full account of the *conative* basis of the agent's reasoning.[284] When the argument is presented in a purely external and assertorical form, the necessity inherent in the PPA's prudential reasoning is neglected. However, once it is recognised that Gewirth's argument proceeds *from the perspective of the PPA's interest* in pursuing their purposes, the PPA's claim to a prudential right follows directly from their concern to ensure that other PPAs must not be allowed to deprive them of the goods needed for such purposive action. MacIntyre objects that the necessity of such prudence need not entail any concept of a right, which he construes as only intelligible to agents educated in a particular social and historical milieu.[285] However, Gewirth correctly points out that despite different historical and cultural descriptions of the concept, political systems for the protection of agents' basic interests are *universally* prescribed and this *is* essentially what the concept of such a right entails.[286]

In this regard, as Beyleveld notes, it is crucial to distinguish between claiming rights in a prescriptive sense which is justifiable on the basis of the agent's prudential interests, as opposed to a "recognitional" sense, whereby the right is effectively legitimised only by the approval of others.[287] As Gewirth's argument is only aimed at justifying the PPA's *prudential right* that others must not deprive them of the goods needed to pursue their purposes, the legitimacy of that right is not subject to the approval of others. Rather, it is the *agent's* implicit recognition of those indispensable goods which validates the inference from (3) My freedom and well-being are necessary goods (for my agency) to (4) I have rights to freedom and well-being. Therefore, the historical fact that agents' interests have long been protected by some implicit conception of rights is not even required to establish the validity of Gewirth's argument in any case.

Nevertheless, it is worth noting that the appeal of MacIntyre's criticisms stem from a common assumption that any concept of individual rights is necessarily confined to societies which share the cultural traditions of the Enlightenment. These traditions are characterised as primarily valuing the individual's freedom of thought and expression, in contrast to those cultures which emphasise the merits of some form of collectivism. Yet, these crude distinctions are largely at odds with the actual anthropological and psychological data collected from many historical periods and locations. For example, prior to any contact with 'individualistic' European cultures, the egalitarianism of foraging tribes was not maintained in the absence of any concept of the individual's distinct contribution to the common good, as Gewirth himself notes in reference to the reciprocal *entitlements* recognised by individuals in the most primitive social systems.[288] Gewirth also cites further substantial evidence from "ancient Greece, Rome, and the middle ages"[289] to counter MacIntyre's assertion that concepts of

entitlement or right "have not existed universally in human societies".[290] As shown in Chapter Two, the very survival of such a 'collectivist' culture is dependent upon each *individual's* capacity to *value* the contributions of *other individuals*, thus reciprocally safeguarding their own welfare at the same time. By the same token, even in the midst of the most excessive promotion of competitive individualism, there is no shortage of mutually beneficial collective enterprises that are voluntarily conducted.[291] As Gewirth points out, "even if it is recognised that individuals are strongly influenced and perhaps even constituted to some degree by the communities to which they belong…in the final analysis, it is individuals who act."[292] Furthermore, as shown by the cross-cultural evidence in Chapters Two and Four, it is only through the *agency of individuals themselves* that social rules and institutions are constructed and *reconstructed* "as when slaves revolt against their masters, or in other revolutionary situations. Hence, it is unduly conservative to presuppose that rights-claims must presuppose social rules or institutions which already exist."[293] If that were the case, *new* ideas or values would never emerge from existing social systems, which is evidently *not* the case.

MacIntyre also objects to this attempt to derive the concept of a right from an agent's evaluation of some purpose as "good", asserting that the latter concept has no deontic implications.[294] Yet, it must surely be conceded that "a necessary condition of any person's claiming a right to anything X is that X seems to him to be good. He need not think it is morally good, or good in terms of his own self-interest, but he must think that it is directly or indirectly good according to whatever criterion he accepts in the given situation."[295] Gewirth correctly points out that in this sense, "a right claim is correlative with and logically equivalent to a strict 'ought'-judgment that other persons ought at least to refrain from interfering with the agent's freedom and well-being."[296] Of course, this logical equivalence does not suffice to authorise any claim to goods which are *not* essential for maintaining prospective agency needs. But it does suffice to justify the agent's prudential prescription that others ought not to prevent those needs from being met. The deontic justification to assert that need as a right becomes clear if it is alternatively conceived as issuing a merely contingent or optional obligation. In that case, in denying that all other persons should refrain from interfering with his agency needs, the agent effectively "accepts that it is permissible that other persons interfere with or remove his freedom and well-being". Yet, as Gewirth notes, such a concession *contradicts* the agent's own recognition that, whatever their purposes, freedom and well-being are *necessary* goods for their agency.[297]

With regard to the other primary objection against Gewirth's derivation, while conceding the validity of the PPA's prudential requirement to claiming the goods needed for their continuing agency, Hare argues that "the universalisation of this claim would only yield the claim that there is a prudential requirement on other similar agents to seek the necessary conditions for achieving *their* purposes."[298] However, Beyleveld makes the crucial point that Hare has neglected to apply the principle of universalisability from the *internal* perspective of the agent making the prudential right claim. From that perspective, the principle of non-contradiction necessarily entails that "where I consider – indeed, am required to consider – that I have a right *by virtue of being a PPA*, this "by virtue of" being a sufficient condition, I cannot consistently refuse to grant this same right to PPAO." (PPAO =other PPAs).[299] Thus, even though Gewirth's Principle of Generic Consistency cannot establish *all-purpose* categorical rights on such grounds, the principle of universalisability certainly does show that specific

moral rights to freedom and well-being are directly deducible on the basis of each PPA's prudential reasoning, in accordance with their conative nature.

In Gewirth's argument, prospective agency is assumed to have an unquestionable value for an agent who may nevertheless come to have extreme doubts that "*E* is good". Indeed, on Gewirth's own account, once a PPA loses interest in any purposive activity, such a person can no longer rightfully claim those very goods that Gewirth recognises to be necessary if agency is to be maintained. Ironically, those persons who are most in need of the goods that would enable them to become prospective agents are thus deprived of the right to reclaim them. Consider a case of the dedicated pianist whose hands become paralyzed as the result of some accident. Suddenly deprived of the capacity to perform the one activity that had genuine value for him, he is at a loss to find value in any other activity or project sufficient to make his life worth continuing. Although such obsessive single-mindedness represents an extreme case, many people with a typically wider range of interests will, at some time in their lives, experience some loss or disappointment so profound that it causes their entire value system to collapse, so that no good prospects are conceivable. While few people reach such a state of depression that would prevent them from envisaging any worthwhile action in future, it is quite common at such times to cast doubt upon the value of one's own well-being, so that even eating and sleeping become unpleasant tasks. In eventually succumbing to extreme hunger pangs or fatigue, do we act as purposive agents?

To the extent that food and sleep are biological necessities that are not consciously chosen, it would appear that one does not need to reflect upon their value in order to act upon them. Thus, it may be argued that the act of responding to these physical pressures does not involve purposive agency. The event of *falling* asleep, in particular, would not seem to be precipitated by any intentional control, in the way that *going* to sleep implies. Although, as shown earlier, most organisms require some degree of autonomy in order to regulate and adapt their responses to environmental events, they need not be aware that they possess that capacity and consequently lack the deliberative power to control its functioning. Yet this lack of insight into the possibility of taking effective action, or of exercising self-control is symptomatic of a person in a depressive mental state. In such a state of self-denial, the agent can hardly regard the act of eating as an autonomous choice reflecting an interest in self-preservation.

Subsequently, upon recovering a sufficient capacity for self-reflection and a renewed interest in at least one activity or goal that is prized enough, the agent will surely then recognise the value of having the freedom and well-being to succeed for that purpose. But this only shows that it is still the current value assigned to the specific purpose(s) envisaged that justifies the claim to those goods needed to exercise prospective agency, rather than a general all-purpose necessity. Furthermore, following Williams' important psychological distinction noted earlier, there is normally a threshold in agents' positive evaluation of their projects that make some "unconditionally" prized, while others are more easily surrendered. When we regard our commitment to certain projects as unconditional, we desire to live for the sake of their successful completion. As long as agents are so highly committed to such goals, the right to the means for continuing agency is then rationally required, as their necessary conditions. However, the evaluation of such commitments as "unconditional" does not thereby justify the claim that freedom and well-being are unconditionally required to realise them. The internal

strength of one's commitment to an end has no fixed bearing upon the external means that are necessary and sufficient to pursue it. It is the specific nature of the activity that sets the degree of freedom and well-being that must be minimally required for its success.

By their nature, those projects to which we are most committed often require a greater investment of time and resources that would certainly imply a right to the continual need to exercise personal freedom, so that any unforeseen obstacles or newly discovered means can be readily negotiated and explored. It is this voluntary, *dispositional* power of freedom that Gewirth claims to be the generically necessary condition for *prospective* agency, in distinction from its necessity for engaging in particular, occurrent activities.[300] Thus, the particular judgment "*E* is good" is not only intended to reflect the agent's current evaluation of some immediate end-in-view, but should also be regarded in terms of the agent's prospective capacity to pursue a range of other possible ends. As Gewirth notes, the agent's recognition of autonomy as an indispensable condition for purposive action is clearly demonstrated by the determined resistance given against any attempt to suppress it, even when not currently engaged by any particular purpose that would require such self-control.[301] Nevertheless, the strength of this concern for personal autonomy will directly depend upon the strength of agents' commitments to their projects. In turn, that commitment will be influenced by the availability of means to pursue them successfully without too much stress.

If sustained action on behalf of cherished goals is continually frustrated, or expectations are progressively eroded by poor rewards, the power of autonomous agency may appear to be so ineffective that the agent may come to judge it a worthless illusion that is utterly dispensable. On the other hand, the personal responsibility for making good choices may seem too costly for the person to bear, so that autonomy assumes a negative value for prospective agency. In both these circumstances, such persons may still value certain purposive activities to some degree, while having an indifferent or even hostile regard for their own autonomy. In this respect, Regan offers an imaginary example of "Zeke, a worshipper of Baal, whose sole purpose is to glorify Baal by performing rituals Baal has commanded."[302] In this case, Zeke need only value his freedom to the extent that it is required to perform these religious rituals. Likewise, his well-being need only be sustained to meet the physical and mental demands that such rituals involve.

To render this example more psychologically plausible, it must be admitted that even the most extreme religious zealots are likely to be concerned with more than one particular purpose. But in cases where religious demands are all-consuming, governing every aspect of the individual's thought and action, personal autonomy has no value beyond that of allowing the disciple to observe these beliefs and practices by following established rules. It is not uncommon for members of strict religious cults, or extremist political organisations to commit themselves to the ideals of the group to such an extent that their own personal autonomy and well-being is radically identified with that of the group, as expressed in ritualistic slogans like "freedom or death!".

When the totality of an individual's personal goals comes to be dictated by the goals of some other person or group, it may well be argued that prospective, purposive agency has evidently been surrendered in any case. As Gewirth notes, those who willingly choose to reject making decisions on their own behalf, still make that choice in their full capacity as agents and continue to be "at least

prospective agents, so long as it remains under their dispositional control whether they will resume occurrent control of their behaviour."[303] Yet, by conceding that agency is thereby subject to the *condition* that a person has such a disposition to effectively control their choice of purposive actions, Gewirth admits that it then becomes necessary to examine how such conditions can best be established and maintained, so that agents are not permanently incapacitated by lacking such dispositional control.[304] Furthermore, Gewirth has still not succeeded in demonstrating that agents must conceive the preconditions for their purposive actions in generic-dispositional terms. Following Gewirth's own analysis of the reasoning process, it is only as a direct result of *first* having some purpose in mind that the agent *then* has a disposition to direct his actions toward it. Therefore, it is the occurrent expectation that "*E* is good" which elicits the disposition to value the freedom and well-being that is needed to realise that end.

In Regan's example, as Zeke is only interested in obeying Baal's commands, he will not have reason to be disposed to valuing any activity which does not contribute to fulfilling his religious duties. Thus, far from having a generic concern for any purposes he might have if he were later to lose his faith in Baal, Zeke's prospective agency is inseparably linked to his continuing commitment to Baal. Being a zealot worth his (pillar of) salt, Zeke could not conceive of the possibility that Baal might have feet of clay, which will shortly be resting on a beach in the Bahamas, with Zeke's life savings also resting in Baal's Swiss bank account. But when Zeke is eventually forced to confront the evidence for such deception, it will equally serve to expose the fallibility of his own judgement and the consequent dangers of committing his freedom and well-being to a singular purpose whose value proved to be unfounded. Even if Zeke's faith is never disrupted by doubt and Baal continues to provide all the goods required for his disciples' ritual activities, the possibility that such faith may easily be misplaced should be reason enough to require that agents always retain the power to reassess their actions in the light of unimagined consequences.

Unfortunately, the rationality of such prudence is characteristically lacking in the judgement of agents when they commit themselves wholeheartedly to ends that will henceforth restrict the range of their conceivable actions. In the case of the religious zealot whose actions are confined to the strict performance of duties determined by dogmatic rules and principles, the natural process of evaluating ends in *conditional* terms has been subverted, if not arrested entirely. This is aptly described and evident in the 'brainwashing' exhibited in the behaviour and thought processes of members of some religious cults, who do indeed seem to have lost the dispositional control that would allow them the freedom to reassess their goals and conceive alternative possibilities. In Gewirth's terms, their occurrent freedom is fully constrained by the dispositional control of an external source of absolute value. In effect, these persons are no longer agents acting on behalf of their own prospective interests, with their unreflective 'autonomy' reduced to choosing the appropriate predetermined response to a fixed range of stimuli. Here, the influence of environmental conditions can be seen as a crucial factor in reducing the value of full prospective agency. By isolating themselves from the wider community and creating an environment designed to limit the range of stimuli to which they are exposed, it is hardly surprising that the members of such cults should gradually lose the disposition to regard their agency as having any value outside that limited environment.

In making a lifelong commitment to restrict their agency to the performance of a predictable range of activities in a cloistered environment, such persons need only claim those goods as are necessary to sustain that commitment. While Zeke may well be criticized for being imprudent in committing all of his resources to the production of ends that are of dubious benefit to enhancing his own future agency, he cannot be charged with any *logical* error in denying the value of whatever means might be needed to pursue other ends that he does *not* value in the least. Thus, Zeke's consuming belief that "Baal-worship is good" only warrants him valuing the means for conducting Baal-worship. As Regan notes, it is precisely because Zeke has this devotion to Baal that he cannot value having the freedom to devote himself to the rituals of worshipping Marduk.[305] In this respect, if Zeke cannot imagine himself *ever* being disposed to valuing the ideals espoused by Marduk's followers, nor any other likely alternatives, these possibilities are effectively ruled out of future consideration. As such, Zeke's commitment to Baal cannot be regarded as an irrational denial of his prospective agency, as such commitment still entails the implicit acknowledgement of the freedom and well-being necessary to maintain that disposition i.e. to keep faith with his religious beliefs.

The above criticisms of Gewirth's conception of generic necessity limit the extent and legitimacy of a agents' rights to the goods required for their particular purposes. Furthermore, as the categoricity of such rights is obtained on the *hypothetical* condition that the agent maintains such purposive intentions, an agent's categorical right to freedom and well-being is contingent upon their dispositional and prospective interests, which need not always prevail. Thus, in order to avoid injurious lapses in these *evolving* conditions that would lead to the occurrent loss or restriction of their agency rights, from the agent's own prudential perspective it is rational to adopt a strategic and dynamic approach to the protection of their prospective agency. This approach thereby extends Gewirth's original argument, by considering the conditions of agency in terms of Dewey's evolutionary pragmatism and the ontogenetic processes described in Chapter Four. This is illustrated in the following section by contrasting examples in the evolving scope of different agents' autonomy and well-being.

(ii)   Autonomy, Well-Being And Purposive Commitment

Both the dispositional control and concern for future agency that characterise the PPA are as vital to the agent's capacity for commitment as they are for avoiding it. In order to prevent oneself from succumbing to the temptation to abandon an important project when faced with immense challenges to its success, it is resolute commitment that is rationally required. This applies equally to the pragmatic formation of scientific theories which frequently demonstrate their validity in spite of apparent evidence to undermine them.[306] But it is no less a feature of practical reason in any other realm. The dedicated researcher's hunt for a cancer cure, Zeke's search for enlightenment, or even Don Quixote's deluded quest are each motivated by a single-minded commitment to goals whose success often depends upon avoiding the exploration of alternative pathways that appear to offer more immediate solutions but soon prove to be deceptive or unreliable. Therefore, the expression of autonomy is not

only evidenced by the agent's adaptability in selecting alternative means and ends where appropriate, but also proves vital in ensuring that a chosen course of action is successfully completed.

Purposive activity often entails the articulation of fixed means and ends, processed in a rigid sequence that has proved to be optimal for producing the desired result. Applied to such tasks, instrumental reasoning bound by predetermined rules to guide the agent's success will then constitute a more rational approach than the standard theory of rational choice, whereby the best decision is determined simply by weighing the costs and benefits of the current end-in-view. By anticipating and mentally rehearsing the sequence of likely outcomes yielded by different options, the agent can make a "sophisticated" choice that affords much greater control over long-term, global consequences. As the range of consequences considered is not limited to a "myopic" view of short-term actions that highlight only the immediate effects of a particular decision, sophisticated planning allows agents to take precautions to ensure that their most valued goals will not be obstructed by a lack of opportunities, or corrupted by one or other means taken to realize them.[307]

In decision theory, the mythical story of "Ulysses and the Sirens" presents a paradigm case for the superiority of this *dynamic* rationality, as described by McClennen. Informed by the experiences of earlier travellers who were unable to resist the Sirens' alluring call, Ulysses takes measures to ensure that he will not have the capacity to succumb to the attraction. For the preservation of dispositional autonomy, the rationality of such planning is necessarily dynamic in its assessment of how particular, occurrent choices impact upon the progressive achievement of overarching goals. As such, it accords well with Frankfurt's hierarchical conception of preferences by which individuals identify themselves at various levels of commitment. On that account, a genuine expression of the agent's freedom of will consists in its conformity to the person's second-order volitions. Each higher order preference can essentially be defined as a second-order volition to identify and act upon that preference, so that even if the highest preference may mark the end of a long series of higher order desires and volitions, a decisive commitment to that final preference effectively terminates the series.[308] Thus, the extreme measures taken by Ulysses reflect the seriousness of concern that motivates him to direct the course of his journey, as far as the gods will allow. Likewise, the unwilling addict is opposed to the occurrent desires which constrain or nullify his disposition to act upon that higher order preference. A naïve first-order account of the psychology of rational choice may indeed explain the addict's original decision to experience the immediate pleasure produced by the drug, even when weighed against the risk of eventual addiction. But as a strategic account of rationality, it offers only a one-dimensional, immature model of reasoning that neglects the authentic agent's need to exercise judgement and control over those ends that are most highly valued.

Gewirth's dialectical construction of the PPA's evaluative judgement is focused only on the isolated end-in-view, ignoring the strategic dimensions of reasoning in the broader social context. In this respect, his conception of the autonomous agent is more akin to that depicted in Sartre's early version of existentialism, where the individual's choices are abstracted from any consideration of their causal interactions with others' material choices and goals. Even in their 'concrete' relations, others are represented primarily as *obstacles* to the successful realization of the individual's projects, rather than basic elements in their composition and completion.[309] If human reasoning were to proceed in such a

myopic fashion, interactions would never be able to achieve the lasting stability evident in all manner of social customs and conventions. However, the evolution of strategic reasoning is necessarily concerned with preserving and enhancing the conditions for *successful* action in the agent's *actual environment*. Thus, Ulysses' strategy of having himself tied to the ship's mast succeeds because it is especially adapted to resisting an identifiable threat in a particular, known environment. But there is nothing unique or rare in such thinking. On the contrary, the need to adapt one's responses to anticipating the actions of others is commonplace in navigating any social environment.

From daily interactions with family and friends in the home, to travelling alone in a completely foreign land, the agent's actions are typically informed and motivated by normative expectations of others. Those expectations hence become a crucial factor influencing the range of ends an individual can hope to achieve in a given environment. Those ends will tend to be ranked accordingly, based on each one's probability of success, as judged from the agent's present perspective. But as one's personal perspective is bound to be subjected to challenges presented by the competing interests of other agents, their likely intentions and reactions must be given significant weight in the process of evaluating one's own projects. However, this consideration of social costs and benefits that is a normal feature in the psychology of motivation, plays no part in Gewirth's analysis of the rational structure of the PPA's grounds for acknowledging the necessary conditions of their actual purposes.

Consistent with his attempt to demonstrate the PGC as a purely formal constraint for deducing the universal conditions necessary for *any* purposive action, Gewirth cannot allow the contextual content of a specific value judgement to have any power in warranting the agent's *generic* rights to freedom and well-being. Still, as Jan Narveson correctly points out, in claiming rights to freedom and well-being, "we will buy the package of rights that will do best for us, in view of the various ends that we have or can expect to have, and not those that some moralist may think we *should* have."[310] Those like Zeke, who have few aspirations, will have little to gain from subscribing to a universal system of extensive positive rights that would require them to make onerous contributions to the welfare of all human beings, in return for the minimal rights to subsistence required for their own private pursuits. If Zeke's claim to freedom and well-being is genuinely aimed at protecting his perceived interests, as Gewirth agrees it should, then his dominant belief that "Baal-worship is good" would hardly lead him to grant every other person the positive right to the goods needed for whatever purposes imaginable. From Zeke's perspective upon the wider society, not all human purposes are worthy of support, while others are actually likely to threaten his own livelihood.

Similarly, the Palestinian will not see any benefit in granting Israeli settlers the right to deprive him of the land he presently occupies. He will recognise his future interests to be best promoted by restricting those rights to his compatriots. His strategic decision will depend on the extent to which he perceives his prospects to be threatened by his neighbours' competing interests. If his interests appear to be irreconcilable with his neighbours', or if he regards their stance as intransigent, he will need to be all the more vigilant and resolute in defending the right to pursue his own ends without interference. Therefore, he will benefit most by claiming only those rights specifically designed to preserve his basic liberties i.e. to inherited land rights, cultural or religious autonomy, rather than agreeing to support a system of State-imposed rights that extends to his enemies the right to make

claims that diminish *his* prospects. Of course, by rejecting the latter system he also forgoes the opportunity of receiving more assistance for himself and his compatriots. However, as those additional positive rights to receive assistance are gained only at the expense of having to equally *render* assistance for his enemies' activities, then those potential gains are easily consumed in the continual struggle to resist the settlers' encroachments upon his scope for self-determination.

Similarly, the unwilling drug addict's most rational strategy will be one that is most likely to ensure that he maintains the resolve to free his will from compulsion by an unwanted desire. In all such cases where the prospects for continuing agency are at stake, such persons cannot afford to take an *ad hoc*, incremental approach that requires a fresh evaluation of their position at every step. Such an approach is not only inefficient and inconsistent with the dynamic intentionality that is required to plan commitments in the first place. The continual reassessment of progress that guides each movement toward the next set of possibilities only exposes the agent to an increasing risk of making decisions based on new experiences that may seem more worthy of commitment than the original objective. In other words, any number of different 'sirens' may come beckoning with such unanticipated frequency that the agent comes to lose interest or faith in the value of the prior commitment. Alternatively, the myopic agent may even be paralysed by indecision, given their inability to foresee the dangers that lie beyond the immediate end-in-view.

In McClennen's game-theoretic analysis, it is the separability of those immediate choices that threatens the sustainability of a commitment task. Even if one takes the sophisticated approach of envisioning the long-range sequence of options that are opened or closed by taking a particular action, the plan that is most feasible on that basis must still be pursued incrementally. Although the sophisticated planner is able to plot a course that is optimal for ensuring that no foreseeable consequences can arise to render the precommitment unrealizable, it still requires the agent to reevaluate each prospective choice when confronted with the task of acting upon it.[311] In doing so, the agent's disposition to pursue the original course of action is continually threatened by the need to evaluate each new alternative that comes into view. Even if the sophisticated planner has accurately anticipated the full range of available options at every point in advance, there can be no guarantee that one or more of those alternatives might not appear more worthy of pursuit when actually encountered, thereby recommending abandonment of the original goal as the most rational choice.

In such circumstances, where no unforeseen obstacles or lack of information can be cited as valid grounds for reconsidering the feasibility of the project, its successful realization will be largely determined by strength of will. Accordingly, the most rational strategy for any particular agent concerned to ensure that they should not be deterred from their commitment, will be one that guarantees that their future choices will accord with their current preferences, thereby removing the threat of any subsequent *akrasia*.[312] As McClennen points out, this resolute approach to fulfilling one's intentions is much more efficient than those like Ulysses', which work by imposing costly external restraints upon the power of future choice, effectively undermining one's autonomy.[313] Certainly, insofar as Ulysses' strategy is voluntary and informed by the intention to preserve his autonomy and well-being against the external threat of the Sirens' song, he succeeds in this. However, it is important to note that while his original plan to have himself forcibly restrained is *conceived* in his capacity as a

prospective, purposive agent, this is achieved at the cost of *suspending* that same capacity. Thus, in addition to the economic costs spent on devising and executing such strategies,[314] there is the much greater danger that while thus incapacitated, one will be powerless to take action against any other threats that may appear with little warning. For instance, in the case of Ulysses and his crew, while passing the island of the Sirens they will have to be ready to fight a sudden storm or repel an ambush by pirates, requiring the full use of all eyes, ears and limbs without hesitation.

Ulysses' strategy of enforced commitment is specifically designed to succeed in situations where the agent's own will is likely to be insufficient for the task. As such, it may well be the most rational strategy for the 'unwilling' addict who continually 'wills' to take the drug, in spite of his contrary preference. In such cases, where the occurrent impulse is so strong as to physically overwhelm a person's most authentic preferences, then such persons can no longer be said to "control their behaviour by their unforced choices".[315] Whilst lacking effective internal autonomy, they must rely on external means of ensuring that their most considered, long-range preferences will prevail. As the Ulysses story shows, this will almost inevitably include the intervention of other agents who can effectively act on behalf of another's autonomy, by setting up conditions to forcibly restrict that person's capacity to act upon impulses which defeat their true preferences. Just as Ulysses relies on his crewmen not to untie him until the danger has passed, the unwilling addict relies on his carers by trusting that they will not respond to his subsequent demands for the drug, no matter how insistent. Of course, there is always the possibility that a hitherto akratic person may actually change their true preferences, even after instructing others to treat such appeals as symptomatic of their weakness. The unwilling addict who, after a number of failed attempts to quit, has insisted his friends promise to keep him locked in his room for several days, may find the suffering so unbearable that he now *genuinely* comes to prefer a life of continued addiction. In that case, as long as the change is not based on self-deception induced by the current withdrawal symptoms, his friends now unwittingly cease to act on behalf of his true preferences.

Notwithstanding such complications, it can be seen that, firstly, the conditions of agency cannot be adequately construed as generic-dispositional goods that are independent of a person's occurrent evaluations of their projects. Secondly, the above examples only lend further weight to the earlier analysis of the pragmatism of reflective thought, wherein the evaluation and implementation of personal projects are themselves evidently subject to the conditional support of other agents, in accordance with dialectical principles of mutual selection. Precisely because such conditions are *evolutionary*, freedom and well-being cannot be considered as powers independent of the actual conditions in which they are manifested. Rather, it is only in the social and economic dimensions that inform the individual's purposes that freedom and well-being have their function as necessary goods.

Therefore, the individual agent does not simply value these goods as indispensable for his own ends, with no regard for the presence of other agents acting in pursuit of *their* own ends. This failure to adequately acknowledge or anticipate the intentions of others is characteristic of the condition known as autism, which severely diminishes the power to plan and negotiate any goal that involves the interests of others. Likewise, as shown in the discussion of Hawks and Doves, even the most selfish individuals cannot succeed in their ends without taking others' ends into account. In short, even if only

for the purpose of preventing others from interfering with the pursuit of his purposes, the PPA's social environment will necessarily constitute a limiting factor in the rational assessment of the means/ends he devises.

Gewirth rejects Narveson's contention that the PPA is rationally required to evaluate the costs and benefits of his purposes, on the grounds that this utility criterion is based on prudential self-interest, which is only one of many contingently variable principles that agents might follow.[316] Gewirth contrasts these utilities with "the rationally preferential status and the qualitative preconditions of any agent's acting to obtain any *other* utilities and benefits."[317] In addition to the voluntary power of freedom, these qualitative preconditions are: (1) the basic goods which "comprise certain physical and psychological dispositions ranging from life and physical integrity (including such of their means as food, clothing and shelter) to mental equilibrium and a feeling of confidence as to the general possibility of attaining one's goals."[318] But this is not enough, as the agent must also necessarily value retaining (2) "nonsubtractive" goods, which consists in those required to maintain his existing "level of purpose-fulfillment", as well as obtaining (3) "additive" goods that contribute to an increase in that level.[319]

While recognising Gewirth's rationale for distinguishing levels of well-being by their "degrees of necessity"[320] for purposive action, Narveson notes that the content of both additive and non-subtractive classes of goods will vary depending upon the purposes for which they are needed.[321] In this respect, Narveson's criticism stems from the same observation made in Regan's subsequent article, already examined above. But Narveson is especially interested in stressing the *consequential* considerations implied by the agent's prudential outlook, which necessarily weighs the value of any purported goods in terms of their contribution to satisfying that agent's particular preferences. Furthermore, as prospective, purposive agents are most concerned with preserving their future prospects for continuing agency, they will be most rational by currently attempting to maximise the opportunities for obtaining goods sufficient to satisfy these future preferences. Accordingly, all three levels of well-being will indeed be objectively required in order to secure these long-term interests. However, such concerns only give the rational agent all the more reason to weigh the costs and benefits of attempting to retain sufficient nonsubtractive goods or to acquire such additive goods likely to be needed for future goals. Therefore, from the perspective of the individual agent, the right to claim a certain level of well-being is not simply a matter of deducing the universal conditions required to justify *any* agent's actions, but will be conditioned by each agent's assessment of the prospects for satisfying *their own* preferences, given the extent to which the prevailing preferences of others are likely to inhibit or enhance these.

Consider a typical scenario involving a more common and less severe addiction than that faced by Frankfurt's unwilling drug addict. A heavy smoker who desperately wishes to quit need not suffer the extreme physical and mental withdrawal symptoms that a frequent heroin user would have to bear. Nevertheless, many an 'unwilling smoker' often finds the prospect of quitting to be a near impossible task, given the relatively low cost, availability and legal sanctioning of their habit. Although most urban dwellers are now much less tolerant of the habit in the interests of their own good health, the smoker is still granted sufficient 'freedom' to smoke in private or segregated areas. If he is to

succeed in quitting, the unwilling smoker will need to take strategic account of these economic and social factors governing his environment, before beginning the withdrawal process. Recalling Dewey's analysis, reflective thinking is initially prompted to find cues in the local environment that can suggest possible solutions to the problem at hand. These suggestions also serve to identify the scope of the problem more clearly and concretely, by immediately drawing attention to the range of possible causes of the problem, as well as possible solutions. In this case, these preliminary investigations will lead the smoker to conceptualise his task as ideally demanding the extinction of the stimulus to smoke, thereby proposing the removal of the external means of addiction as a leading hypothesis. But when this proposal to avoid exposure to cigarettes is tested - either overtly, or in the imagination – it reveals the strength of the smoker's own impulse to continue the habit.

Once the task of quitting has been clearly defined by identifying the main obstacles to overcoming the addiction, probable solutions will focus upon finding the most practicable means of combating these. Devising plans that are geared to suppressing the urge to smoke are unlikely to succeed if no attempt is made to assess the risks engendered by exposing that impulse to an environment conducive to smoking. Likewise, plans that only seek to avoid possible exposure to such environments are also flawed, in failing to assess the agent's own motivational capacity to resist the internal impetus that may persist, even in the absence of an objective external stimulus. Habit formation provides the perfect illustration of the dialectical selection process that induces adherence to principles or rules by testing their viability against each other in various circumstances. Regan considers the hypothetical situation of an aspiring ballerina who also happens to enjoy smoking.[322] Recognising the habit as an impediment to the progress of her career, she resolves to quit. In that case her decision is clearly the result of weighing the costs and benefits of two competing preferences that cannot both be satisfied without diminishing the rewards of each. While the judgment against smoking is certainly in the interests of her prospective well-being, it is made only in order to realise a specific purpose.[323]

It may be objected that smoking is known to be a risk to general health, or likely to shorten one's life span and would therefore threaten the prospects for pursuing any activity. But this only further highlights the empirical basis for making risk assessments, as opposed to deducing the generic grounds for undertaking any action, regardless of the risks involved. For example, if the unwilling smoker's habit has been nurtured and reinforced by an established career in rock music, with most of his earnings dependent on regularly playing in smoky venues, his decision to quit has direct financial implications which must be weighed against the reduced, but still significant risks associated with passive smoking. These considerations will themselves be evaluated in the light of additional factors that influence the level of commitment to competing means and ends. His age, support and advice from friends, family and health professionals (advice which may conflict among these sources), previous attempts to quit, alternative career options, pleasure in performing etc. are only some of many issues that would play a substantive role in determining the most desirable course of action.

In terms of Gewirth's classification of goods by their degrees of necessity, it can now be seen that an agent does *not* value these "irrespective of the more particular contents he may assign to various of his purposes."[324] On the contrary, it is the particular content of his purposes that requires him to

make a rational assessment of how each can be afforded, given the resources available to him, in the present and future. Therefore, contrary to Gewirth's claim, the need to estimate the value of continuing to pursue a current activity, or of taking on additional projects, is not, in fact, derived from any contingent criterion of value that is secondary to the essential conditions for purposive action. Given that any particular end requires a certain sufficiency of means, in both qualitative and quantitative dimensions, these are the variable preconditions that rational agents must acknowledge as essential to achieving their ends. These economic principles constrain the choice of achievable means and ends but also thereby *enable* the selection of those that are most practicable given the agent's resources. Unlike Zeke, agents are not typically inclined to dedicate their actions to one all-consuming activity, but prefer to explore a range of possible spheres of action comprising many different goals over a lifetime.

Though there may be an unlimited number of purposes that an agent might conceivably imagine to be good, there are obvious physical and psychological constraints upon the quality and quantity of goods that an agent can experience fully at any one time. Each must compete for a sufficient investment of time and energy to make the experience both possible and worthwhile. For these economic reasons alone, ends and means are continually subject to the selection pressures imposed by the agent's current level of well-being. As a result, at any given time the agent will only be capable of acting upon a set of occurrent ends-in-view where each is ranked according to its prospects for finding some degree of completion in the immediate future. However, such a ranking need not follow any strict method of utility calculation in assessing the value of undertaking a particular task. Nor should the agent be expected to value different purposes for the same universal quality, or by the same measure of quantity, as if all purposes required the use of identical resources to realise them with equal efficiency.

Therefore, the "economic rationality" applied to estimating the costs and benefits of proposed actions does not imply that every PPA must be guided by the same principles. If that were the case, there would be little or no disagreement among agents in judging the order in which a given set of goals should be prioritised. In fact, the same person may simultaneously prioritise different means and ends by divergent principles or heuristics that are appropriate for certain domains but not others. For example, in shopping for food a poor student will be economically rational by following a heuristic that ranks their choice of products by price above all,[325] with little value placed on distinctions in quality. However, that same person may direct the progress of their studies by contrary principles that recommend diligent and painstaking research for many hours each day. So there need be no inconsistency or conflict where an agent applies opposing rules to guide tasks in different domains. As in this example, the same person's endorsement of contrasting principles of economic value is best understood as reflecting distinctions regarding the value of the activity itself. The student invests a maximum amount of time and energy in his studies and a minimum at the supermarket. It is his *comparative* expenditure of resources that is governed by economic rationality, regardless of the different means of measuring goods.

Considering that for most agents the judgment '*E* is good' refers to only one of a set of ends-in-view that can possibly be accommodated at one time, its actualisation must come at the expense of the remaining ends. Even the archetypical ascetic or a monomaniacal creature like Zeke must divide his

time into periods for concentrating on routine and ritualised tasks, if only to avoid reflection upon the possibilities relinquished by their lifestyles.[326] Therefore, determining the necessary goods for purposive action is tantamount to determining how prevailing resources can be efficiently apportioned as means toward the completion of each end. The PPA's concern to preserve sufficient goods for further ends requires that any particular occurrent end be assessed for its effect in diminishing the existing level of goods. Gewirth agrees that these non-subtractive goods and whatever additive goods can be accrued must be valued in terms of loss and gain, regardless of the different criteria by which they can be measured as such.[327] But he fails to appreciate how these evolving levels impact upon the selection of occurrent purposes and how those purposes in turn shape the future conditions for purposive action.

As Narveson observes, "an agent who changes his purposes will at time $t°$ regard the nonlowering or raising of his level of fulfillment of $E°$ as good, but at t^1 regard these conditions as evil."[328] It is only while the habit is continued voluntarily that the addict necessarily values having the resources to support it. Seeking rehabilitation, the unwilling addict will now value conditions conducive to the development of behaviours that are fundamentally *opposed* to conditions facilitating the former addiction. In the case of the smoker, at $t°$ the immediate costs of the habit can be readily quantified in the affordable price of cigarettes purchased, while the more distant health risks will typically be discounted or deferred from consideration lest they spoil the puffing pleasure. But once the balance comes to shift to the point where the true costs are more keenly understood, much of the pleasures derived from the habit begin to pale in comparison, so that the reformed smoker at t^1 will now regard 'smoker-friendly' conditions as positively inimical to his well-being.

The extent to which those conditions would reduce his level of purpose fulfillment will depend upon specifics of the social environments that he must negotiate in order to maintain or improve that level. The rock musician whose earnings are mostly derived from regularly working in such smoker-friendly zones is bound to suffer significant costs from his decision to quit. Even if he is able to tolerate exposure to smoke in the interests of keeping his preferred occupation and the lifestyle it affords, that constant effort is likely to have implications even at the level of those basic goods concerned with maintaining the agent's physical and psychological powers. In this respect, the capacity to secure sufficient "food, clothing and shelter" may seem to constitute minimal conditions for the preservation of agency itself, regardless of the activities they support. However, from the perspective of individual agents, even the basic necessities of life are negotiable by their 'degrees of necessity' in sustaining or advancing one set of interests over another.[329] In fact, it is precisely in the interests of his physical health that the ageing rocker would be highly motivated to cease performing, in spite of the risk of impending unemployment. Likewise, the student may choose to subsist on the cheapest available food and accommodation to ensure that sufficient funds are available for course materials and fees. In the interests of living an unencumbered, itinerant lifestyle, some will be happy to sleep under the stars, even at the cost of constantly scrounging for the next meal. Whatever the case, the goods needed to sustain activities in one realm, thereby reduce the prospects of fulfilment in others. As a result of the severe disappointment or regret that follows from having to abandon or indefinitely postpone the pursuit of a cherished goal, in the interests of satisfying other, more urgent demands, the

agent is also liable to suffer a *loss* in the "confidence as to the general possibility of attaining one's goals". As observed in discussing the ontogenetic development of phenotypic plasticity in Chapter One, the autonomy and well-being of most organisms is continually subject to such economic trade-offs. Accordingly, if agents are to minimise or compensate for such losses, any assessment of the necessary conditions for successful agency must therefore give due consideration to such evolving constraints.

As these considerations indicate, in adapting to the conditions of daily life, the process of determining the necessary and sufficient means of purposive action *in general* does not even apply to the pragmatic task of evaluating the costs and benefits of *specific* actions envisaged. The aforementioned loss of confidence occurs in recognition of a lost opportunity to achieve a goal with a specific content. Indeed, it is only because the idea has a certain content that the costs of its realisation can be rationally assessed. For example, the dedicated pianist who loses the use of his hands will have cause to lose confidence in the "general possibility" of succeeding at alternative ends that involve the efficient use of hands and fingers. It is important to note that this general assessment of his prospects is not directed at any generic capacity for agency, but is derived from and restricted to purposes that are *generally dependent* upon the *particular* physical capacity he now lacks. Nevertheless, if the loss of function should severely threaten the competent performance of an activity to which the person is devoted, then disappointment can certainly turn to despair at the possibility of enjoying success in other areas as well. Thus incapacitated, the piano virtuoso who cares only to play literally finds himself 'at a loss' in conceiving new possibilities worth attempting. The acute awareness of his physical paralysis is then manifested in a paralysis of the imagination, where alternative possibilities are barely entertained. But when his agency is thereby reduced, it is not as a result of a rational assessment of his true potential. On the contrary, his defeatist disposition is irrational in its determination to infer general negative consequences from a loss that imposes limitations only in certain areas of action, which have not been adequately explored by the usual process of reflective thought.

In this example, the individual has scarcely ventured upon the initial step in the Deweyan model, wherein the agent is motivated to begin the search for possible solutions to overcoming the obstacle in his path. Having so thoroughly committed his physical and mental well-being to the perfection of a singular pursuit, the possibility of embarking on a new career path is not even conceived as an option. Hence, he is unwilling to circumvent the problem by abandoning his ideal activity to settle for some achievable but unrewarding alternative. Consequently, he perceives his physical incapacity as an insurmountable obstacle to his future agency itself, whose well-being is valued only insofar as it is needed to sustain those skills which can no longer be practised. His predicament highlights the inherent dangers of exhausting one's entire resources in the exploration of a singular source of value, in the presumption that other sources are bound to be unfulfilling by comparison. However, as the sudden paralysis of his hands would rightly be judged an unlikely event, his calculation of the risks to his career would certainly be rational. His worthy regard for the level of commitment required to perfect his skills is not in question, but the fact that this regard comes at the expense of a narrow-minded *disregard* for the comparable value that may also be found in many other pursuits. The value in any conceivable action cannot be adequately judged without considering what

unanticipated effects it might yield when undertaken within different contexts, or employed as a means to different ends. Of course, the economic and social restrictions on readily accessible means and practicable ends will be a significant factor in defining both the scope and need for such experimentation, before making the most suitable career choice. But then, economic and prudential rationality at least will have guided the agent to prioritise a *range* of affordable means and ends in the light of *foreseeable* consequences, so that a strong commitment to any single activity should not jeopardise the chances for fulfilment by other means if necessary.

With this in mind, Regan considers another Baal worshipper, Zelda, who is a more reflective believer willing to concede the possibility that her present belief in Baal may be misguided.[330] Acknowledging the fallibility of human judgment, Zelda wants to retain the freedom to worship some other deity whose truth she may not grasp at present. Yet, in spite of this conscientious concern to ensure that she makes the right commitment, the grounds for her prospective conversion to another faith would be subject to the same fallibility that undermines the earlier belief in Baal.[331] It is important to note that this problem of continual fallibility is not confined to questions of religious belief that are by nature, less dependent upon rational assessment or empirical support.[332] In any realm of action, there is always some probability of encountering unexpected events, or discovering unanticipated implications that can render one's current beliefs suspect, in spite of the best attempts to ensure that all relevant factors have been taken into account. Through social interaction alone, a change in preferences may gradually evolve, as the agent is exposed to a range of perspectives, or "forms of life" not previously contemplated or experienced directly.[333] Indeed, such changes in the focus of concern are indicative of the growth of agency itself, as shown in Kohlberg's description of the progressive stages of moral development.[334] By the same token, such development cannot be sustained by constantly changing preferences in a vain attempt to adapt to every other point of view. Witness the chameleon-like character Leonard Zelig, whose complete lack of self-assurance causes him to readily assume the character and appearance of those in his present company.[335] In his uncritical absorption and imitation of their behaviour, he finds the company of Adolf Hitler and Charlie Chaplin equally absorbing.

The rationality of most prospective agents has naturally evolved to a normative standard that is adapted to the task of making decisions under conditions of prevailing uncertainty. For this reason, Zeke's blind commitment to maintaining his present concerns and Zelig's continual submission to changing conditions are extremes that are only beneficial in the absence of these uncertainties. If Zelig could be absolutely certain that the world was populated only by persons concerned to care for each other's well-being, his trusting acceptance would not be pathological or dangerously naïve. If Zeke could only be certain that Baal's one-way trip to the Bahamas was for the purpose of retiring to a life of contemplation, he would not be risking his life savings. By contrast, Zelda's concern for the preservation of her future interests is not dictated by the need to maintain her current convictions, as if she were infallible. On the other hand, being so mindful of this fallibility is liable to perpetually undermine the basis for a firm commitment to *any* beliefs, so that sustained effort and concerted action becomes untenable, resulting in the demise of effective agency. However, while prudence ordinarily directs agents to guard against the possibility of being harmfully misguided by their beliefs or values, that same interest in self-preservation also provides the motivation to *trust* in their ability to

successfully monitor the consequences of their actions, in order to avoid costly disappointment or regret, but without having to forgo the benefits that can only accrue with sustained commitment.

As the Ulysses example demonstrates, prudence will sometimes recommend enforced commitment to measures specially designed to protect the agent against threats that are almost certain to occur. Fully expecting that his immediate preferences will change in response to the Sirens' song, he takes prior actions to ensure that he will nevertheless continue to pursue his original quest. Knowing that his future self at t^1 will suddenly prefer to cease doing what he now most prefers at t^0, his aim is to secure the long term interests of his present self. As McClennen argues, given that resolute commitment here amounts to having the present self "tyrannize" over one or other future selves, "what possible *rational* ground could your future self have for accepting such a regimen?"[336] Of course, this is precisely why it then becomes necessary to resort to enforcement, as some later self will see little reason to maintain a commitment to preferences that it does not now hold. So it is that Zelda would not want to allow her present commitment to Baal the power of effectively preventing her from worshipping some other deities in future, especially if she has good reason to expect that these later preferences are likely to be significantly more informed by further experience and evaluation than is available to her present self.

Although her later preferences are likely to be more considered, this only provides further reason to perpetually postpone any commitment, as each self at any given time will not be as wise as any subsequent self. Thus, even if at t^0 Zelda is committed to Baal, though willing to concede the possibility of converting to Marduk worship in future, the eventual realization of that possibility at t^1 does not provide sufficient grounds for her to dismiss the further possibility of becoming disillusioned with Marduk at t^2 and returning to Baal worship all the wiser, yet none the wiser in relation to still later preferences. Here, the concern that her present beliefs should not be allowed to hinder the satisfaction of her future goals only leads her to change her preferences too frequently and too late, so that in this case commitments are rendered futile under the tyranny of *future possibilities*. However, given the social and economic forces that restrict the range and degree of satisfaction that can actually be obtained in the pursuit of any future preferences, the mere possibility that Zelda might prefer to worship Marduk at some unforeseen date need not prevent her from remaining fully committed to Baal for the present, on the condition that continuing investment in Baal worship should not impose feasibility constraints or costly burdens on defection. For example, if admission into the Baal cult should be obtainable only at the cost of transferring her life savings into Baal's coffers, or if Baal's teachings are premised upon the belief that Marduk is the 'Evil One' representing everything that Baal abhors, such demands would indeed constitute serious pragmatic reasons for withholding commitment. Furthermore, for the reasonable and reflective agent that Zelda is intended to represent, such costly conditions would not only raise immediate suspicions about the true worth of Baal, but would also prompt her to proceed with caution before making any irretrievable commitment of resources that could be much better allocated towards deep and long-lasting interests which have already stood the test of time.

Therefore, notwithstanding the possibility of serious misjudgment or self-deception, it is reasonably safe to assume that Zelda would not be easily seduced into worshipping a god with dubious

credentials or onerous conditions attached. By her concern for preserving the freedom and well-being needed to satisfy her future preferences, she imposes her own economic conditions on the costs of her present commitments. But if her future preferences as yet had no conceivable content or probability of fulfilment that could be measured against those currently held, she would have no means of estimating the viability and worth of her current commitments. Thus, it is only because she already has sufficient knowledge of the conditions involved in Marduk worship that she is able to anticipate both the costs and especially the benefits that might eventually lead her to change her priorities. If she were merely indifferent to those possible benefits, it would not be a factor worthy of any consideration in evaluating the status of her present commitment to Baal.

By the same reckoning that motivates her *not* to exclude Marduk worship from among her future options, Zelda cannot afford to be indifferent to certain other possibilities that would be unthinkable if not for the need to recognise their endorsement by others. As Regan suggests, Zelda's willingness to admit the possibility of conversion to Marduk should not be thought to imply an attitude of global openness towards all other deities, as being more or less preferred as possible candidates for future worship, for there are bound to be gods like Dagon, who might idealise or condone values or practices that Zelda could only regard with outright disgust or contempt.[337] Thus, if Dagon were to represent ideas that are fundamentally antithetical to those espoused by both Baal and Marduk, or he perhaps even called for the destruction of all rival religions and their 'infidel' followers, Zelda would find absolutely no value in creating conditions that would enable her to worship such a god in future, if she were somehow seduced into converting. On the contrary, like Ulysses, she would want to protect her freedom and well-being *against the possibility* of being used in the service of ends perceived as inimical to her future agency itself. As such, her antipathy towards certain beliefs or practices is not some irrational prejudice that signifies a failure to appreciate diverse cultural perspectives. It indicates the robust, rational coherence that governs the dialectical selection and prioritisation of her prevailing system of values, wherein the *prescription* of a chosen set of preferences necessarily implies the *proscription* of opposing sets. This is this logical demand for consistency that rational agents must observe in determining the necessary conditions to fulfil their purposes, rather than the generic consistency that Gewirth insists upon.

This analysis and modification of Gewirth's argument completes Part Two, and furnishes the basis for the theme of Part Three, which focuses upon the rational and pragmatic constraints which are necessary and sufficient to preserve and enhance the agent's autonomy. Ultimately, as this shared capacity enables and qualifies rational agents to select and adopt the principles which will best protect and advance their autonomy and well-being, that enhanced autonomy thereby increases the scope and practicality of future rational choices. Accordingly, the concept of an organic social contract can be succinctly characterised in terms of this synergistic relationship between *the autonomy of rational choice* and the *rational choice of autonomy*.

---

[263] Gewirth, *Reason and Morality*, pp.37-39

[264] Ibid., pp.102-104 & p.133

[265] Ibid., p.135

[266] Frankfurt, op. cit., pp.17-18

[267] Kant, Critique of Pure Reason, B117-B118

[268] Kant, *Critique of Practical Reason*, 3-5 & 49-51

[269] Kant, *Groundwork of the Metaphysic of Morals*, 439

[270] Gewirth, *Reason and Morality*, p.62

[271] Ibid., p.44

[272] Ibid., p.45

[273] Gewirth, "The "Is-Ought" Problem Resolved," in *Human Rights: Essays on Justification and Applications* (University of Chicago Press, Chicago & London, 1982), p.116

[274] Kant, *Groundwork of the Metaphysic of Morals*, 442-443

[275] Ibid., 444

[276] Ibid., 412; In regarding all such human "inclinations" as "impulsions", Kant notes that "by acting in accordance with his natural constitution…it is nature which would make the law. This law, as a law of nature, not only must be known and proved by experience and therefore is in itself contingent and consequently unfitted to serve as an apodeictic rule of action such as a moral rule must be, but it is always *merely heteronomy of the will*: the will does not give itself the law, but an alien impulsion does so through the medium of the subject's own nature as tuned for its reception." 444; Therefore, any "special predisposition of humanity…certain feelings and propensities", in short "everything that is empirical" must be ruled out as the basis for any moral principle, so that "the principle of action is free from all influence by contingent grounds, the only kind that experience can supply." 425-426

[277] Deryck Beyleveld, "Gewirth and Kant on Justifying the Supreme Principle of Morality," in *Gewirth : Critical Essays on Action, Rationality, and Community*, ed. M. Boylan (Lanham, MD, Rowman & Littlefield, 1999) p.111

[278] Donald Regan, "Gewirth on Necessary Goods: What Is the Agent Committed to Valuing?" in Boylan ed., G*ewirth*, p.46

[279] Ibid., p.55

[280] Ibid.

[281] Ibid., p.48

[282] François Marie Arouet de Voltaire, *The History Of Candide: Or, All For The Best* (London, Simpkin Marshall, n. d.)

[283] It may be objected that a categorical right cannot be founded upon a hypothetical condition. Of course, in terms of a simple distinction between any *contingent* categorical and hypothetical imperatives, this is indeed the case. Otherwise the distinction would make no sense. However, as freedom and well-being constitute *necessary* goods for *any* purposive action, they must be *categorically* valued (on pain of contradiction), but only *if* the agent has some purpose(s) to pursue. Therefore, the categorical requirement comes into force only under the latter *proviso*. By contrast, as already noted in Chapter Three, a hypothetical imperative to drink if one is thirsty is not by itself sufficient grounds to warrant a categorical response i.e. to drink *regardless* of its possible deleterious effects upon one's prospective agency.

[284] Michael Boylan, *Gewirth*, pp.2-3

[285] Alisdair MacIntyre, *After Virtue: A Study in Moral Theory* (Notre Dame, Indiana; University of Notre Dame Press, 1981), pp.64-65, cited in Deryck Beyleveld, *The Dialectical Necessity of Morality : An Analysis and Defense of Alan Gewirth's Argument to the Principle Of Generic Consistency* (Chicago, University Of Chicago Press, 1991), p.153

[286] Beyleveld, op. cit., pp.155-156

[287] Ibid., pp.147-148

[288] Gewirth, *Reason and Morality*, p.101

[289] Alan Gewirth, "Rights and Virtues," *Review of Metaphysics* 38 (June 1985), p.747. See *Reason and Morality*, pp.100-101 & pp.372-373, notes 19-26 for these references

[290] MacIntyre, *After Virtue*, p.67

[291] For many examples of such diversity within modern cultures, see the following articles in Larry P. Nucci, Geoffrey B. Saxe & Elliot Turiel (eds.) *Culture, Thought and Development* (Mahwah, NJ & London, Lawrence Erlbaum Associates, 2000): Claudia Strauss, "The Culture Concept and the Individualism-Collectivism Debate: Dominant and Alternative Attributions for Class in the United States," pp.109-114; Larry Nucci & Elliot Turiel, "The Moral and the Personal: Sources of Social Conflicts," pp.115-137; and Michael J. Chandler, Christopher E. Lalonde & Bryan W. Sokol, "Continuities of Selfhood in the Face of Radical Developmental and Cultural Change," pp.65-84; see also the following articles in *Morality in Everyday Life*: Cecilia Wainryb & Elliot Turiel, "Diversity in Social Development," pp.283-313; and Hart et al, "Moral Commitment in Inner-City Adolescents," pp.317-341

[292] Alan Gewirth, "Is Cultural Pluralism Relevant to Moral Knowledge?" *Social Philosophy and Policy* 11, no.1 (1994), p.34

[293] Gewirth, "Rights and Virtues," p.747

[294] MacIntyre, op.cit., pp.64-65, cited in Beyleveld, op. cit., pp.231

[295] Gewirth, *Reason and Morality*, p.76

[296] Ibid., p.78; derivative support for this logical equivalence can be found in Hare's analysis of the analogous equivalence in the prescriptive use of 'good' and 'right' as an adjective. See Part 3 of *The Language of Morals*, esp. Ch.12

[297] Gewirth, *Reason and Morality*, p.80

[298] R. M. Hare, "Do Agents Have to be Moralists?" in *Gewirth's Ethical Rationalism: Critical Essays with a Reply by Alan Gewirth*, ed. E. Regis Jr. (Chicago, University of Chicago Press, 1984), p.55

[299] Beyleveld, *The Dialectical Necessity of Morality*, p.270

[300] Gewirth, *Reason and Morality*, pp.52-53

[301] Ibid., p.52

[302] Regan, op. cit., p.48

[303] Gewirth, *Reason and Morality*, p.53

[304] Ibid., pp.266 & 292

[305] Regan, op. cit., pp.53-54

[306] Waller, op. cit., p.20

[307] Edward F. McClennen, "Rationality and Rules," in *Modeling Rationality, Morality, and Evolution*, ed. P. Danielson (Oxford, Oxford University Press, 1998) p.20

[308] Frankfurt, op. cit., pp.20-22

[309] Jean-Paul Sartre, *Being and Nothingnes: A Phenomenological Essay on Ontology*, trans. H. Barnes (New York, Washington Square Press, 1966), pp.360-2 & esp.Part 3, Ch.3. See also Christina Howells, *Sartre: The Necessity of Freedom* (Cambridge, Cambridge University Press, 1988), pp.19-21

[310] Jan Narveson, "Alan Gewirth's Foundationalism and the Well-Being State," in *Journal of Value Inquiry*, no.31, p.490

[311] McClennen, "Rationality and Rules," p.22

[312] Ibid., p.23

[313] Ibid., p.24

[314] Ibid.

[315] Gewirth, *Reason and Morality*, p.59

[316] Ibid., p.57; See also Alan Gewirth, "Replies to My Critics," in *Gewirth's Ethical Rationalism*, p.231

[317] Gewirth, "Replies to My Critics," p.230

[318] Gewirth, *Reason and Morality*, p.54

[319] Ibid., pp.54-56

[320] Ibid., pp.62-63

[321] Jan Narveson, "Negative and Positive Rights in Gewirth's *Reason and Morality*," in *Gewirth's Ethical Rationalism*, p.101

[322] Regan, op. cit., p.64

[323] Ibid.

[324] Gewirth, *Reason and Morality*, p.61

[325] Lawrence Haworth, *Autonomy: An Essay in Philosophical Psychology and Ethics* (New Haven, Yale University Press, 1986), p.94

[326] Gewirth, *Reason and Morality*, p.59

[327] Ibid., pp.55 & 57

[328] Narveson, op. cit., p.101

[329] Ibid.

[330] Regan, op. cit., p.66

[331] Ibid., p.67

[332] Ibid., p.66

[333] Cf. Wittgenstein's use of the term "forms of life" to denote rule-governed activities which can only be understood in terms of their normative function in a particular social context - See *Philosophical Investigations* (Oxford, Basil Blackwell, 1968), § 23

[334] Lawrence Kohlberg, *The Philosophy of Moral Development: Moral Stages and the Idea of Justice* (San Francisco; London, Harper & Row, 1981), pp.17-20

[335] Woody Allen, *Zelig*, Orion Pictures & Warner Bros., © 1983

[336] McClennen, "Rationality and Rules,"p.23

[337] Regan, op. cit., p.66

CHAPTER SIX

PART THREE:

THE AUTONOMY OF RATIONAL CHOICE AND THE RATIONAL CHOICE OF AUTONOMY

CHAPTER SIX

PRINCIPLES AND METHODS OF RATIONAL COHERENCE

From the strategic and dynamic perspective described in the previous chapter, it is now possible to consider the basic contractual principles and methods which can justifiably serve to guide the choice of both rational and reasonable moral and political principles designed to protect the evolutionary conditions of agency. The first section defends a more extensive application of Gewirth's argument for consistency in accordance with the logic of non-contradiction and universalisation, endorsing the validity of Hare's account, as will be demonstrated with examples.

(i)   Logical and epistemic consistency: Universalisability

Gewirth rightly cites the principle of non-contradiction as the one formal axiom that ultimately defines the structure of consistency in practical and moral reasoning. To reiterate: the individual PPA cannot consistently claim that certain goods are necessary to achieve his purposes, whilst at the same time denying that all other PPAs should need such goods. If the possession of properties X & Y (freedom & well-being) constitute necessary conditions for being Z (PPA), then by definition, *every Z must require X & Y.* This logic thereby reveals another principle of consistency implicit in moral thinking: the notion of universalisability, which Gewirth takes as a criterion of sufficient reason for the ascription of universal and categorical rights to all PPAs.[338] However, as the above examples indicate, while every Z does indeed require X and Y, they do so by varying 'degrees of necessity', depending on the range and content of their purposes. The preferential selection of their purposes ensures that those most preferred can only be pursued by denying the capacity to act upon contradictory preferences. In effect, as a matter of consistency, the affirmation of some well-defined positive value implies the concurrent rejection of other values which would negate that potential.

To illustrate this process, imagine the distinct value systems that might plausibly represent the prevailing concerns of each of three hypothetical character types as models of evolving agency. Situated in a contemporary social environment, the strategic outlook of each of the 'Z's cited above can be seen to typify different levels of effective agency. Thus, Zeke's unquestioning religious zeal represents a cognitive approach that applies to any set of ideas to which a person is dogmatically committed, with no respect for the validity of other alternatives. Zelig's perpetual distrust of his own

judgement leads to the opposite extreme of avoiding the responsibility of commitment, while Zelda's prudential commitment represents the normative level of self-concern exhibited by the PPA.

Firstly, consider the religious zealot whose actions are guided by what he claims to be the strict observance of commandments or rites to honour his God, whose will is revealed in oral testimony or scriptures held to be sacred. Unperturbed at the prospect of restricting his agency toward the fulfilment of a single all-consuming purpose of dubious reward, he resolves to give his full commitment to God by becoming a monk. Nevertheless, his commitment may still be the end result of the process of pragmatic reflection outlined earlier. Assuming he has given full and serious attention to the foreseeable costs as well the expected benefits of an ascetic life, his commitment would be rational. As long as he has at least *considered* the possibility that his belief may be misplaced, or even that some other god might be more worthy of respect than he presently imagines, those possibilities will have entered into the mix of prospective ideas that marks the early stages in the reflective process. Therefore, even though they may be short-lived suggestions that are quickly dismissed from thought, in favour of contemplating the rewards of the afterlife that his own god promises, all of these rejected ideas continue to function in the form of *disvalues* that set conditions on what other values can be rationally accepted or rejected concurrently, while also limiting the scope for any coherent change of preferences.

Following Dewey's analysis, it was noted that reflection is typically initiated by the urgency in finding a viable solution to a problem that impedes progress toward some end-in-view. Psychologically, religious commitment can be seen as the solution to a lack of meaningful purpose, where the individual is perplexed by uncertainty about the direction of his life. William James observed that all religious belief evolves in response to a sense of "uneasiness", which "reduced to its simplest terms, is a sense that there is *something wrong about us* as we naturally stand", so that religion comes as the solution in that "*we are saved from the wrongness* by making proper connection with the higher powers."[339] Thus, the religious impulse is initially provoked on an emotional level to counteract the generalised discomfort or existential 'dread' that follows from some serious doubt about the true nature of one's place in the universe. A particular religion will offer a solution that is able to resolve these doubts by proposing the existence of a higher realm or force which transcends the material world. On a personal level, the individual's discomfort is transformed into joy as the worldly ends-in-view that were earlier regarded with great uncertainty, now appear imbued with a higher purpose ordained by God. It is commitment to this higher *self* that ironically reveals the self-interested motive at the heart of religious belief, in spite of the zealot's demonstrative acts of self-denial. As James notes "the religious individual tells you that the divine meets him on the basis of his personal concerns".[340]

Here then, a question of potential inconsistency poses a threat to the intelligibility of the zealot's belief system. If he is to maintain the firm belief that his asceticism truly expresses a humble commitment to his god's wishes, he cannot afford to admit the contradictory belief that he is only acting to secure his personal salvation, as this would only revive the sense of uneasiness and doubt that his religion had successfully dispelled. Yet, neither can he afford to abandon his belief in the reward of personal salvation, as this is indeed his ultimate concern. Unable to dispense with either belief, the contradiction is resolved by reconciling the believer's personal interests with God's commands, so that

the two magically coincide. The zealot is thus able to convince himself that his personal salvation is *desired by God*, first and foremost, thereby allowing him to suppress any consciousness of the self-interest that has driven his commitment from the beginning. This explains why the zealot must insist upon such supererogatory demonstrations of faith, exalting the glory of God at every opportunity to reinforce the belief that he does not act for his own sake. Thus, the zealot's belief that God "meets him on the basis of his personal concerns" is infused with an entirely new meaning that identifies *God's* concern for him, rather than the zealot's own self-interest as the reason for worshipping Him.

Such a case would seem to suggest that some semblance of rational coherence can be achieved by a simple process of self-deception or wishful thinking, affirming only those beliefs and values that are not in immediate contradiction with our actions. A common facility for rationalising inconsistencies in our behaviour has been well documented in psychological experiments, where individuals are faced with the problem of accounting for contrary beliefs or attitudes that generate "cognitive dissonance" among one or more of them.[341] For instance, faced with strong evidence that smoking causes cancer, a smoker becomes aware that his behaviour contradicts his desire to avoid getting cancer. Self-justification is attempted by various strategies that aim to reduce the dissonance by modifying one or other beliefs or actions to achieve a reasonably coherent fit. So the smoker might reconstitute his thinking via some: (a) change of belief (e.g. not convinced that smoking causes lung cancer) (b) change in behaviour (e.g. quitting smoking) (c) reevaluation of the contradictory behaviour to minimise the significance of the conflict (e.g. by noting that he does not smoke much, or only smokes "socially" at parties etc.)[342]

As this example shows, rational coherence is not only governed by strictly logical principles such as the law of non-contradiction, but also involves a wider recognition of epistemic and pragmatic inconsistencies that are not captured in purely deductive inferences. The agent's capacity to change beliefs, values and actions which would come into conflict with his preferred purposes implies the recognition of antagonistic relations between properties or potentials that, while not logically impossible, are nonetheless mutually incompatible. A certain level of practical experience and observation of facts and probabilities is a prerequisite for being able to anticipate and adjust one's beliefs and behaviour, so that they are adequately adapted to the mutual limitations of opposing forces. For instance, imagine a person who has not yet learnt that water has an antagonistic effect upon fire. Suddenly he sees his house in flames yet makes no attempt to throw water upon it, even though it is freely at hand. He cannot be charged with any logical contradiction, as would be the case if he claimed to be a bachelor while married and knowing that 'all bachelors are unmarried'. Nevertheless, his inability to infer the negative effect of water on fire has the most serious consequences for the maintenance of his freedom and well-being. But once he understands the variable strengths of these elements in different environments or when applied to different objects, he is then able to harness their mutual antagonism to his advantage. He now learns how to reverse the balance of power by using fire to boil water, enabling him to further enhance his well-being by cooking his food.

The importance of learning the principles that govern the behaviour of fire or water in close proximity to each other, contained within different environments composed of other competing elements, applies equally to the discovery of the evolutionary principles that govern the adaptive fitness

of mental and behavioural phenomena. Once it is recognised that rationality itself is an adaptive process whereby beliefs, ideas and values are made coherent by the practical results of their continual interaction, the law of non-contradiction alone scarcely furnishes the ground for a principle of *sufficient* reason in justifying the use of certain means for certain ends. While it is a sufficient criterion for falsifying particular beliefs or actions as concurrent determinants of an agent's actions - as in the paradigm case of the married man's claim to be a bachelor – it cannot measure the degrees of necessity among values that fluctuate in accordance with changing conditions encountered by the agent. A man who at twenty years of age claims to be a confirmed bachelor does not contradict himself if he eventually marries at the age of thirty, for the simple reason that his character and value system has most likely evolved in response to the value of new experiences which could not be fully appreciated at that earlier age. Indeed, if reminded of his earlier belief, he need not even attempt to defend its rationality from his present, contrary perspective and may coherently dismiss it as belonging to a naïve, younger self. However, even in the unlikely event that his preferences have shifted so dramatically that he now subscribes to a set of beliefs or values diametrically opposed to those of his younger self, the transition in values can be effected without contradiction. He may well agree with Kierkegaard's ethical polemicist Judge William in proclaiming:[343]

> I am not a young fanatic who tries to put forward his theories; I am a married man, and I certainly dare to let my wife hear that all love in comparison with repentance is but children's babbling. Nevertheless I know that I am a good husband, I, who even as a married man am still struggling under the triumphant banner of first love.

This honest acknowledgment of the personal struggle to reconcile conflicting ideals perfectly captures the distinction between the fanatic's bad faith in denying any prevailing contradictions in his attitude, in contrast to the willing self-reflection that is a prerequisite for rational coherence in the mature agent's value system. As an example of the former, Kierkegaard notes the contradictions in the mystic's zeal for loving God which is shown only by rejecting the actual conditions of his earthly existence, thereby contravening God's will in placing him there. Even worse, the mystic may perceive himself to have a special relationship with God due to his possession of some personal trait, such as humility, that God especially prefer.[344] Such monumental self-deception is required to mask the equally monumental self-absorption that lies behind it. For this reason, all expressions of fanaticism, dogmatism and the like betray a state of arrested development that is usually associated with a sense of self yet to achieve the stage of critical self-reflection practised by mature agents.

In terms of the agent's capacity for autonomy, Haworth cites four basic stages that are identifiable from a consensus of empirical research in developmental psychology. [345] Stages one and two mark the transition from the child's increasing competence in exercising self-control to the point where the conscience begins to form, though its directives are still only the internalised beliefs of parents or other authority figures and are not subjected to any further reflection. It is only at the third stage where the normal capacity for autonomy is evident in the disposition to critically reflect upon received ideas and values, even if only intermittently maintained. The final stage is characterised by the

achievement of complete independence from both internal and external constraints so that the agent approaches a level of "unrestricted critical competence". While the final stage represents an ideal attainable only under similarly ideal conditions where agents need no longer compete for resources, such complete independence from social and economic constraints is impractical and arguably undesirable for beings who have evolved strong social instincts. For precisely the same reason, even within the parameters of the normal capacity for autonomy, individual agents cannot afford to be continually engaged in critically scrutinizing their own beliefs and values. Until their concrete effects are observed, they remain only hypothetical candidates for consideration in motivating purposive actions. Critical reflection upon one's beliefs, values or ideals arises in order to resolve some doubt about the results of acting to realise one or more of them. Once those doubts have been resolved by successfully meeting the agent's expectations, they need not be called into question until further conflict or doubt arises.

By contrast, as Hare points out, what characterises the typical though "impure" fanatic is his persistent refusal or inability to engage in critical reflection on occasions when it is unequivocally called for by the destructive consequences that it entails.[346] The "pure" fanatic, such as the Nazi who would not hesitate to follow the ruthless logic of his own self-destruction upon discovering both his parents to be Jews, would rarely if ever be encountered given the costs of maintaining such convictions.[347] He cannot be accused of inconsistency in his attitude and the risk that he may become a victim of his own harmful recommendations is in fact minimised by his willingness to reflect upon their universal application. However, like the pure Hawk discussed earlier, the pure fanatic can only survive in a society with a large population of victims to prey upon. To maintain the source of his fanaticism, he not only requires the continued presence of those who are the target of his prejudices, but he also depends upon the presence of a significant proportion of persons naïve enough to be seduced by his rhetoric. In effect, like a vampire the pure fanatic creates his own unwitting army of impure fanatics though not by the universal logic that he himself accepts. Agents with the normal capacity for critical reflection would be inclined to recoil from their prejudices after grasping the unpalatable moral consequences they entail. While normal agents are still prone to episodic displays of fanaticism, these are likely to be the result of misinformation or short-lived lapses in moral thinking, rather than any incapacity or reluctance to engage in critical reflection. As such, they are generally well-disposed to resist appeals to dangerous excesses in thought and action. Unfortunately, however, as even the most recent historical events have shown, in many populations there is no shortage of those impure fanatics who have never succeeded in developing a moral conscience that transcends the internalisation of authoritarian norms, as defined by the second stage of Haworth's synopsis.

From Regan's analysis, the particularity of an agent's purposes was shown to circumscribe the degree to which that agent must be committed in valuing the freedom and well-being that is necessary to achieve those purposes. Furthermore, Gewirth's own recognition of the degrees of necessity by which different classes of goods contribute to effective agency might seem to deny the validity of deriving any universal application from a singular judgement that "*E* is good". Any particular *E* need only express a contingent preference alterable by further assessment of the costs and benefits involved in its potential satisfaction over other competing preferences. However, as argued earlier, this

contingency should not be interpreted as an indication that individual preferences are bound to be arbitrary expressions of desire with no rational basis for adjudicating among them. On the contrary, it is the capacity to effectively regulate the expression of even the most primitive first-order desire that is a condition for the development of normal agency. Thus, even the strictly biological desire of hunger is routinely subject to rational control by the strategic selection of means designed to prevent or reduce its arousal (e.g. regular mealtimes), or by acting upon alternative preferences that occurrently inhibit the desire (e.g. smoking). Nevertheless, the decision to assign a particular weight to one preference over another cannot be fully determined by such processes, as there is no independent criterion for measuring the value of each.

While the consequences that result from acting upon the chosen preference in various situations does provide a measure for determining the scope and efficacy of its realisation, that knowledge alone cannot suffice to allow the agent to set a fixed value upon the activity. For instance, the fact that Mary now chooses to smoke a cigarette in order to suppress her appetite is an action that is rationally consistent with her further preference to lose weight. Similarly, experimenting with a range of dietary regimens allows her to make a rational choice of their value in contributing to that end. The preference to lose weight thus confers intelligibility upon the concomitant preference for smoking. As an end-in-view, the proximate value anticipated in the achieved weight loss will, in turn, be rationally accountable as means to some further goal, such as the aspiration to become a fashion model. Even if Hume is right in claiming that any particular desire is ultimately traceable to universal origins in the pursuit of pleasure and avoidance of pain, the range of actions permitted by such general principles is so diverse that any particular preference that accords with them will inevitably be insufficiently determined on that basis, so there will always be some element of contingency in the selection of any individual set of preferences. Thus, the brand of cigarettes that Mary currently prefers may simply be the result of the contingent fact that it happens to be the one that is most freely available among her friends or colleagues. Her current satisfaction with brand *X* does not provide sufficient rational grounds for never choosing brand *Y*, any more than her desired weight loss should commit her to smoking as one of the means to achieve that goal. The weight loss, in turn, is no less subject to continual reappraisal in the light of progressive difficulties which may not be adequately appreciable beforehand, so that the target may need to be revised or may yet be abandoned if a life on the catwalk should thereby lose its glamour. Therefore, the contingency lies in the inherent uncertainty of any evaluation, rather than being attributable to some arbitrary quality that is not within the agent's control.

Nevertheless, contingency is inevitable where rational choice is constrained by agents' cognitive or perceptual limitations, making it virtually impossible to discriminate between objects displaying equally relevant similarities, as in the case of Buridan's ass.[348] As Bratman notes, these Buridan cases are as commonplace as the task of choosing a box of breakfast cereal at the supermarket.[349] Applied to the earlier supermarket scenario, imagine that all the Easter eggs on the supermarket shelves are replicas from the same mould and manufacturing process, identical in virtue of size, weight, labelling and expiration date. Faced with such an equal array of items, even the most discerning adult can employ a strategy no more rational than that of the child who simply takes the one nearest in reach. But again, while contingent, the preference is not arbitrary, as it is guided by an

implicit hypothesis that any further examination in the number or precise details of individual items would prove to be a fruitless exercise.

Hare describes moral thinking as a process akin to that of Popper's account of scientific theorising, wherein hypotheses are conceived and tested, not for any demonstrable truth, which can never be conclusively established, but are instead progressively eliminable by generating results which falsify them.[350] While Popper's theory is itself drawn from his interpretation of the evolutionary principles governing scientific progress, it is has been rightly criticised for its erroneous emphasis upon falsification as the major goal of scientific enquiry. Such stringent criteria can only impede progress for both the scientist and moral agent. Unless *inductive* leaps of the imagination are taken beyond the available evidence, neither scientific nor moral progress is possible. But once it is granted that the search for ideas to *guide action* can only proceed via the process of inductive, instrumental reasoning, then both scientific and moral thinking are equally constrained by the same method of enquiry, as Dewey has shown.[351]

As moral principles are constructed as *hypothetical* prescriptions recommending different actions in different contexts, the universalisability of any prescription may therefore seem to be entirely unwarranted, as a law of deduction that has no application to moral reasoning. But this assertion would itself be unwarranted. While hypotheses are formulated and tested by rules of induction, those rules do not succeed by invalidating the laws of deductive logic, as previously noted in regard to the principle of non-contradiction. In considering Hare's application of the term, it is important to avoid interpreting "universal" to mean "general".[352] As induction consists in reasoning from the particular to the general, it may be thought that Hare is employing the term as an inductive rule of inference. If that were the case, then he would certainly be guilty of falsely claiming that the violation of a general moral principle such as "one ought to keep one promises" amounts to a logical *error*.

Rather, the notion of universalisation here applies to the identification of similarities among particular cases, so that what is deemed relevant to prescribe in case *X,* where *a* and *b* are identified as the determining factors for judgement, must also be prescribed in every other instance where those same factors are equally relevant to that situation. For this reason, the universalisability of any particular prescription is not limited by its degree of specificity. As Hare notes, both the general statement that "one ought never to make false statements" and the quite specific version that "one ought never to make false statements to one's wife" are both universal claims.[353] The latter judgement can be seen as a qualification of the former, as a result of the person having been prompted to consider its validity under certain conditions. In effect, the impersonal "one" here is an accurate indication of the fact that it is the relation between husband and wife that is identified as the relevant similarity that *must* be present in *every* instance where such a relation exists. In this respect, the universalisation of a prescription follows from the same principles of universal categorisation in the description of phenomena. Thus, if a person identifies an object as 'red', then all objects similarly perceived, must also be described as 'red'.[354] Although, of course, evaluative terms cannot be said to describe any objective properties as 'good', but only express an individual's judgement of their goodness for some subjective interest, this distinction is of no relevance to the issue of logical consistency in judgement.

Despite the potential latitude in variations among the agent's subjective preferences, those choices *are* constrained by consistency in the agent's judgement of the *objective* qualities perceived as good for the satisfaction of a particular preference. In this respect, Hare notes that even the simple choice of a cup of coffee can be scrutinized as more or less rational in terms of its coherence with both *logic* i.e. in the correct semantic application of the identified concept(s); and *facts* i.e. the expected consequences of the prescribed action.[355] Thus, a person whose confusion of words in a foreign language leads them to accept an offer of whisky instead of coffee has not yet accurately identified the foreign equivalent of 'coffee' to render their decision sufficiently rational. More importantly, however, is the degree of rationality displayed in the person's expectations of the likely effects of their action. Even where a person does understand perfectly well that they are "in Egypt and the coffee is Turkish", or the drink is dispensed by "one of those machines that we have in England, which produce either tea or coffee and it is hard to tell which", or worse still, where the person is unaware that the coffee is strong enough to cause them to suffer a heart attack, Hare notes the irrationality involved in accepting the drink while "not knowing the properties of the coffee" on offer.[356] As such, the rationality of a particular preference is indeed measurable in terms of objective, factual consequences which also imposes constraints of logical consistency. Thus, for example, James' assertion that, in addition to the standard recipe, a good martini must be "shaken not stirred'' commits him to making the same judgement about *every* martini that is shaken so as to produce a *similar* taste, as long as he continues to hold that preference with equal strength.

In spite of the common observation that there is "no accounting for taste", the universalisability of any particular preference does, at least, provide a strictly logical measure of accountability among an individual's concomitant preferences. In the above case, assume that James' strong aversion to stirred martinis has been formed as a result of one or two unpalatable experiences in sampling them for the first time. A number of contingent factors may have contributed to that initial experience, so as to give a false impression of the effects discernible from the act of stirring. For example, as an accompaniment to a hot and spicy meal, or as a beer 'chaser', draughts taken from exactly the same drink may produce remarkably different effects upon the palate. Assuming that such effects would be equally discernible if James is served a *shaken* martini in the same circumstances, then it is these similarities which are relevant in determining his taste for martinis. Of course, as long as he remains unaware that his present aversion is based on false beliefs, he cannot be charged with any inconsistency in maintaining his current preference. Nevertheless, like Hare's example of the ignorant coffee drinker, James' choice of martinis can still be deemed irrational by the extent to which it is held with conviction, in spite of the weak hypothetical support informing the choice. Therefore, insofar as he fails to even *consider* the more *plausible* alternative explanation, his choice is irrational. To remain a rational agent, he could not *ignore* the cumulative weight of this competing evidence, particularly when it becomes glaringly inconsistent with the effects experienced by many others who have consumed martinis in exactly the same circumstances as he. In that event, in his epistemic interests as a rational agent, he will be obliged to test the alternatives himself and accept the hypothesis with the much greater empirical support. He will then be logically obliged to revise his preference so that its universalisation is consistent with its *true* constitution under these alternative hypothetical conditions,

where the act of being "shaken not stirred" is now irrelevant. Accordingly, his new preference will commit him to prescribing that any person with a similar preference for martinis ought to avoid the newly identified conditions which are liable to render the experience equally unpalatable e.g. when consumed with hot, spicy food, or immediately following *x* number of full strength beers etc.

From the logic of universalisation, it follows that rational agents with identical preferences of equal strength are thereby rationally required to take the same action in the same situation, as the prescriptive meaning of 'ought' implies. Of course, it may be objected that no two individuals or situations can be accurately described as identical in every respect. However, this criticism has no bearing upon the process of identifying similarities that are only relevant to *particular preferences*, which are themselves identifiable by the particular *goals* to which they refer. As noted in discussing Dewey's example of the bushwalker, the extent to which similarities are noteworthy is determined by the specific nature and exigencies of the situation, which is itself defined by the relative strength of the person's desire to traverse the gully, in preference to setting up camp for the night. Therefore, universalising the latter preference held with the same degree of commitment would amount to recommending an *equally* urgent search for firewood, for which logs should likewise be chosen for their *similarly* high combustibility etc. As such, all evaluative judgements are universalisable hypothetical imperatives which become subject to logical consistency in their expression as *recommendations* to oneself or others with the same preferences in the same circumstances. Nevertheless, as long as recommendations to oneself in the form of personal ideals do not have any discernible impact upon others' interests, then such actions need not be morally universalisable. In this respect, Hare notes that a man who prefers to take some physical exercise before breakfast is not being inconsistent if he fails to disapprove of his neighbour's contrary preference to stay in bed reading the newspaper. But if the ideal is prescribed to *others* as one that *ought* be pursued, then the universalisation of such a judgement becomes a moral directive, logically committing the "ascetic" to disapproving of his neighbour's behaviour.[357] Yet, (as noted in Chapter Three with the example of the self-destructive person), it must be remembered that this universalisation *only* imposes a logical commitment upon the person who issues the prescription, having no epistemic relevance to those who do *not* share that person's ideals.

Regardless of unique contingencies that may yet distinguish individual cases, it is the relevant similarities in person and situation that grounds the logic of their universalisability. As it follows directly from the deductive principle of consistency in identifying particular items by their possession of the same universal properties, it cannot be claimed that universalisability is an inherently evaluative notion, still less a "substantive moral principle" in the guise of logic, as Mackie suggests.[358] For the same reason, it cannot be argued that evaluative judgements belong to some special faculty of reasoning where rules of consistency can be successfully discarded. In response to the first objection, any suspicion that universalisability is implicitly moral in meaning, framed in terms of fairness, equality or any other substantive notion, would be demonstrable by its favouring the endorsement of prescriptions that accord with such 'moral' values, over those which do not. However, this cannot possibly be the case, as the universalisation of any particular judgement does not, by itself, impose any *a priori* restrictions on its content. It only imposes logical restrictions in the consistency of *relations*

between separate judgements that cannot be maintained *together* without contradiction, as would occur in any statement which asserts that "I ought to act in a certain way but nobody else ought to act in that way in relevantly similar circumstances".[359] As Hare's discussion of the fanatical Nazi illustrates, far from being a moral principle, universalisability alone cannot invalidate commitment to the most extreme prescriptions that would, however, be unacceptable on moral grounds of equality and fairness.

In considering the evolution of basic moral concepts and rules, their emergence was shown as dependent upon primordial categories and modes of practical reason that constitute the evolved 'architecture' of the mind. As the identification of universally relevant similarities in phenomena is essential to the perception and control of their effects, universalisation is itself a *universal a priori* principle of logical categorisation and deduction that thereby sets formal constraints upon the evolution of all subsequent domains of thought. Therefore, the claim that universalisation is essentially a moral concept that has been illegitimately imported into the structure of prescriptive logic is a clear case of "putting the cart before the horse". Furthermore, the supposition that aesthetic or moral systems of thought somehow manage to transcend such logic is clearly untenable, requiring appeal to some superfluous mental faculties whose origins and existence would be very 'queer' indeed.[360]

While his "argument from queerness" is an effective riposte in demystifying claims for moral transcendentalism, Mackie's scepticism towards the notion of universalisability fails to take account of its necessity and logical priority, in spite of his own regard for the significance of evolutionary factors in the formation of ethical concepts.[361] Although admitting that it does have a legitimate function in ethical reasoning, Mackie distinguishes three stages of universalisability, each marked by its progressively weaker logical support. The first stage is defined in precisely those terms described by Hare i.e. implying commitment to the same action in relevantly similar circumstances. Like other critics,[362] Mackie agrees that, so defined, the principle "in some sense, is beyond dispute" in recognising the "irrelevance of numerical differences".[363] However, both the second stage of "putting oneself in the other person's place" and the third which involves "taking account of different tastes and rival ideals" are considered respectively to have diminishing relevance in justifying ethical requirements.[364]

This depiction of universalisability as a process admitting degrees of validity, fundamentally misrepresents Hare's application of its validity in exclusively deductive terms, which accords with only the *first* stage of Mackie's analysis. Mackie's definitions of the second and third stages indicate their deepening level of concern for the interests of others, which he does not consider to follow directly from the "irrelevance of numerical differences" that defines the first stage. Yet, as Singer argues, it is precisely that irrelevance which invalidates any claim that the interests of one individual can take precedence over those of another, given that each person's interests are *equally* subjective, as both Hare and Mackie agree.[365] This irrelevance of subjective distinctions in evaluating one's own and others interests should not, however, be taken to undermine the special value normally assigned to cherished personal ideals and loyalties, such as relations between mother and child, husband and wife etc. As noted above, an individual's own personal ideals need not be subject to any moral constraints unless its pursuit has some negative impact upon the interests of others. Therefore, extending equal consideration to the interests of others need not require one to devalue or demote personal projects that do not have

such moral consequences. Likewise, universalisability does not require a mother to care equally for both her own and her neighbour's baby. As Hare notes, far from excluding the value of such special concerns, universalisability in fact serves to *warrant* their promotion as universal social norms. As Hare notes, all such relations between persons are themselves universalisable *qua* relations so that, in effect, it is perfectly reasonable for a person to defend the priority of those relations by claiming that "I have only such duties to *my* family as I am willing to allow other similar people to have to *their* families."[366] In short, these personal loyalties and concerns are as universalisable as any other preference. Thus, the irrelevance of 'numerical differences' among agents, which requires the equal consideration of affected parties' interests, is not incompatible with agent-relative restrictions on the scope of personal concerns. As the above example indicates, as long as one is consistent in prescribing that others should be allowed to express the *same* preferential support for *their* personal interests, those special relationships can be fully promoted. Furthermore, as shown in Hare's discussion of the pure fanatic, the universalisability of such partial preferences will ultimately tend to promote greater *impartiality* at the political level, as few would be willing to prescribe policies such as nepotism or patriotism to which they themselves may well become victims.[367]

Singer's point can be confirmed in more directly pragmatic terms, by reiterating Dewey's empirical demonstration of the hypothetical nature of evaluative judgements which undermines any attempt to assign a definitive value to any particular preference. Even those moral beliefs which have repeatedly proven to be reliable guides in informing the agent's choice of ends can never be exempted from further assessment, as they are always subject to possible revision in the light of new, unforeseen consequences. Furthermore, as moral beliefs are, by definition, beliefs prescribed to others, they are *necessarily* subject to the logical constraints of universalisability. Therefore, whenever the agent is confronted with a new decision, even in the most familiar and relevantly similar circumstances, it is not possible to imaginatively rehearse all the potential consequences which a particular belief may yield, especially when every belief must continually be validated against other possible alternatives. Thus, in deliberating upon any set of beliefs and values "If *a* conditions *b*, and we are interested in *b*, how can we as rational beings avoid becoming concerned with how *a* affects *b*, and how different forms of *a* condition different varieties of *b*?"[368] In this respect, Dewey observes that even the most profound deliberations of choice "only fixes a disposition which has to be applied in new and unforeseen conditions, re-adapted to future deliberations."[369] Therefore, even "the good of foreseen consequences or of attained consequences is not final or dogmatically determined" as "reflection is instrumental to the creation of *new* consequences and goods when taken in its integrity – or experimentally."[370] Given this continual dependence on reflection, Dewey's approach accords with Gewirth's, in recognising an indispensable value in the individual's freedom of choice as the "capacity for deliberately changing preferences…"[371]

Moreover, neither can there be sufficient grounds to justify drawing a distinction between the act of imagining oneself in another's position and taking account of that person's tastes and ideals. Given that *B*'s preferences are intrinsic to the position that he holds, *A* can scarcely imagine being in *B*'s position *without* holding *B*'s preferences. Therefore, Mackie's three stages in fact represent a *single*

process of universalisation, whereby relevantly similar features of person and situation are thoroughly maintained in the act of considering the interests of all those concerned.

Hare provides a classic example of the universalisability of a particular moral judgement that serves to clarify both the logical consequences and extent of its application. Derived from the biblical parable in *Matthew xviii 23*, Hare considers a case where *A* is in debt to *B* who, in turn, is in debt to *C*, where creditors have recourse to a law of the land allowing them to exact their debts by sending their debtors to prison.[372] In such circumstances, *B*'s consideration of whether or not to avail himself of this law cannot be logically restricted to its effect upon *A* alone, as *B*'s debt to *C* puts *B* in precisely the same indebted position as *A*. *B*'s invocation of the law's rationale against his debtor thereby commits him to accepting its equal application to himself as *C*'s debtor. As long as the state of being in debt is identified as the relevant similarity in *B*'s endorsement of the law, he cannot logically deny that *C* ought to put him in prison without contradicting his own acceptance of that same rationale in *A*'s case. Not surprisingly, *B*'s recognition of these undesirable consequences in his own case gives him cause to reject the proposition that he ought to have *A* imprisoned.

As Hare notes, these observations arise as a direct result of deducing the universal implications of a singular hypothetical prescription.[373] However, in this example, *B*'s subsequent rejection of the prescription is readily explained by his acknowledgement of its *actual* application to him personally, rather than from any hypothetical position. While prudential self-interest is undoubtedly a prime motive for *B* in this case, it could not have arisen as a relevant factor if he had not universalised what was originally only a singular prescription applied to *A*. It is only as a result of having *first* identified the description 'debtor' as the essential factor in justifying his proposed action against *A*, that he is *then* able to recognise himself as fitting that same description. Therefore, by the same reasoning, *anyone* who *would* fit that description must also be deemed worthy of the same treatment, regardless of the potential for any actual, or known creditor to take the same action against *B* personally.[374] Thus, *B*'s prudential motive in declining to endorse *A*'s punishment has absolutely no bearing upon the logical consistency that commits him to prescribing the same treatment in all relevantly similar cases. Consequently, *B*'s recommendation in *A*'s case necessarily extends to cases which have yet to arise. As long as the status of 'debtor' remains *possible* and the law continues to authorise the penalty in question, *B*'s judgement must also apply to hypothetical cases, thereby including anonymous creditors who are powerless to exact the same penalty against him.

The logical inclusion of hypothetical cases is thus entailed by the universalisability of the relevantly similar feature which they share with actual cases. But this does not, of course, require that every single hypothetical possibility should be recognised as such in the imagination. Apart from the fact that this is one hypothesis that is impossible to test, given that any relevant similarity potentially includes any number of persons imaginable prior to their actual existence, universalisability does not in fact demand the omniscience of an archangel. For Hare, the archangel simply represents an ideally insightful perspective which nevertheless exemplifies the importance of considering the full implications of our moral judgements, as far as is humanly possible.[375] To that extent, it will often be more feasible to limit the scope of one's considerations to the interests of those who are more directly

affected by the hypothetical prescription, without doing any injustice to those whose interests might only be affected fortuitously.

As noted earlier with regard to Buridan's cases, the factors that are relevantly similar are themselves determinable by the exigencies of the goal to be realized. In recommending an ideal breakfast cereal for a friend who wishes to lose weight, one would only need to be sufficiently informed of the sugar and fat contents of those brands which the friend could obtain without too much trouble. If the product that best meets these criteria should nevertheless include a new artificial sweetener, in amounts which are later found to be carcinogenic when consumed daily over several years, one could not be accused of neglecting the friend's interests. Likewise, there may be a variety of unknown factors which may undermine the validity of any moral judgement, in spite of the best attempts to consider all manner of contingencies which would have warranted withholding the prescription in certain cases. The debtor example is instructive in this respect, as Hare himself suggests, in raising a distinction in financial resources that could be adduced in support of exempting the imprisonment of those debtors with families entirely dependent upon their earnings.[376] But due consideration of such additional factors is perfectly consistent with the principle of universalisability, as their admission only results in the modification of the initial hypothesis to identify new cases which are relevantly similar on these additional grounds. In the same manner in which the general principle against making false statements is tested and given specific approval when considered in the context of marital relations, in this example the relevant status of 'debtor' is found to be too broad in permitting the same punishment to all debtors, regardless of their capacity to pay or other extenuating circumstances.

While the principle of universalisability has thus been shown to have an indispensable and widespread application in setting constraints upon the validity of practical and moral prescriptions, it does not suffice to prevent choices that are potentially self-destructive, self-defeating, or harmful to others' autonomy and well-being. As such choices are liable to be based on false beliefs, misinformation, or miscalculation, it is vital to allow agent's decisions to be adequately informed of the full range of opportunities and resources afforded by the material and social environment. This next section highlights the need to evaluate the adaptive prospects afforded by the various rules, practices, institutions and resources that have evolved within different social ecologies. These constraints and 'affordances' can then be better adapted to cohere with agents' needs as they evolve in these environments. In evaluating these affordances, agents can have recourse to more general principles of epistemic coherence that can be used to assess the rationality and reasonability of prospective moral and political policies. Thagard's multicoherence approach is thus presented as setting the *general* coherence parameters which would have the widest application to guiding the choice of contractually justified principles and norms.

(ii)   Interpretive Coherence: Constraints and Affordances

The progressive refinement of evaluative judgements is driven by the demands of logical and epistemic consistency described in the preceding discussion. Universalisability and its justification by the principle of non-contradiction provide the most stringent criteria for testing the validity of beliefs

and values solely in terms of their mutual coherence. However, as noted in discussing the prevalence of the psychological phenomenon known as 'cognitive dissonance', laws of deductive inference cannot be employed to combat the type of epistemic inconsistencies based on unwarranted estimates of value or probability. In the example of the smoker confronted with the risk of dying from lung cancer, three typical responses were cited, only one of which qualifies as a sufficiently rational remedy i.e. resolving to quit the habit. Where this response is appropriately rational in its assessment of the high long-term risks associated with smoking, the other responses are strategically aimed at discounting those risks by either finding fault with the evidence, or by self-deceptive appeals to the low frequency of indulging. Not surprisingly, as the term suggests, such rationalisations testify to the strength of habits and passions in making reason their slave, even though Hume wrongly construes this tendency as a normative necessity rather than a psychological tendency that may be rectified.[377]

The extent to which rational methods can be successfully employed to correct such self-serving biases will reflect psychological differences among individuals in their disposition to exercise control over their desires and values. Notwithstanding the influence of these habits, there are a number of common cognitive biases in probabilistic reasoning which have usually been cited as indicating that human reasoning does not always conform to statistical standards, such as those measured by Bayesian estimates. Also, psychological experiments have repeatedly shown that certain elementary heuristics are regularly applied to problems where lack of information reduces the accuracy of probability assessments.[378] Yet, as Gigerenzer has forcefully argued, the supposed illusions and errors that are generated by such habits, in fact present quite reasonable and accurate estimates when the problem to be solved is defined within a specific domain.[379] Given that the human mind evolved to calculate probabilities within localised ecological boundaries, rather than in long-range estimates of unknown frequencies, these inherited heuristics and norms of induction would be especially adapted to interpret and predict events by their local occurrence.

The ecological basis for a rational assessment of probabilities is clearly shown by Gigerenzer's example in a jungle setting, where a parent is faced with the task of choosing whether her child should be allowed to swim in a nearby river, or to climb trees instead. The parent's limited information consists in knowing that there has only been one accident in the river in the last hundred years, where a child was eaten by a crocodile, but that a dozen children have been killed by falling from trees. Then, only the day before, she learns that her neighbour's child has been eaten by a crocodile. In this case, an *unbiased* statistical estimate of the chances that her own child might be killed while swimming is reached by simply updating the base rate of past swimming deaths from one to two in a hundred years, thus barely registering any further cause for concern over the children swimming, in spite of the crocodiles that have perhaps only recently returned to the river.[380]

As this example illustrates, there are clear evolutionary advantages for an organism equipped with heuristics specifically adapted to the task of registering sudden changes in the local environment. The heuristic applied in the above example succeeds by assigning a higher probability to a more recent local event, in spite of its statistical rarity over a long period and unknown probability over wider environments. Therefore, it can be expected to register a significant proportion of false positives. But as

long as it is efficient enough to alert the organism to real dangers on the occasions when these *do* occur, then the information to which they are 'biased' represents the most accurate assessment possible.

In many species, such heuristics function with an immediacy that is more or less determined by the range and frequency of the particular environmental dangers they have evolved to encounter. For example, vervet monkeys in Kenya have evolved distinct alarm calls to alert their companions to the presence of three different predators commonly encountered in that local environment i.e. snakes, leopards and eagles.[381] When these Kenyan vervets migrate into the forests of Cameroon where they encounter hunting dogs for the first time, they immediately adapt their signalling system to identify this new threat by using the leopard call. Unfortunately, despite this instantaneous use of an established heuristic to incorporate new and vital information that can be readily transmitted throughout the troupe, the loud leopard calls only succeed in alerting the hunters to the vervets' presence in the trees, so that they are easily shot. At the same time, however, a subpopulation of vervets in these forests have substantially reduced this danger by modifying the calls to be shorter and quieter.[382]

Although these 'fast and frugal' heuristics evolved to facilitate decision-making and problem solving in conditions of uncertainty, if they are employed routinely or indiscriminately in response to new environmental conditions, then they are liable to prove maladaptive. In deciding that her child should not be allowed to swim in the river, the parent is intuitively guided by a rule that is specifically designed to signal dangers that are proximate in time and space. Even if her neighbour's child had been eaten by a single stray crocodile which had been swept into the river by a rare flood and has since been caught and killed, in the *absence* of such conclusive knowledge the rule succeeds in preventing exposure to a life-threatening risk. But under different conditions where a wealth of reliable information is available and the threat is not so urgent, then the focus upon rare and recent changes is liable to actively increase the probability of dangers being realized, by ignoring or underestimating the totality of statistical evidence that would suggest those changes to be unrepresentative anomalies. It is precisely this rule to assign greater value to local, anecdotal information that enables the smoker to rationalise his habit, by reducing the cognitive dissonance between the value of his desire and the possible threat to his health. Typically, he will recite cases of chain-smoking octogenarian acquaintances who have regularly consumed no less than two packets a day since their teenage years, without suffering so much as a single cough that could not be attributed to some other cause.

As these two examples indicate, the rationality of any *specific* heuristic cannot be judged independently of its application to a *specific* task in a *specific* environment. The full range of ecological conditions in which a rule might continue to function adaptively will depend upon how closely new phenomena can be incorporated into the existing framework without misrepresenting their pragmatic affordances for different tasks. Originally proposed by Gibson, the concept of "affordances" is central to Reed's ecological psychology, in denoting environmental resources or opportunities for action whose availability or scarcity creates "selection pressure on the behaviour of individual organisms; hence, behaviour is regulated with respect to the affordances of the environment for a given animal."[383] It is important to note the generality of the concept in encompassing the phenomenological structure of the environment, so that the *ecology* of such affordances does not exclude the resources and opportunities made available by the actions of other organisms.[384] Thus, the term 'ecology' can be

coherently applied to the *context* of information in any domain. As such, the capacity to interpret information in terms of affordances provides the ecological criterion for making a rational choice between competing heuristics. But as this capacity is itself dependent upon the organism's *prior* capacity to organise information so that affordances can be accurately identified and anticipated, then a method of coherent systematisation of information must be actively and *consistently* maintained by each individual organism. If newly perceived features are vaguely or ambiguously defined when first encountered, this may be a signal of a weakness in the current system of selection, whether it be in the domain of sensory perception or in the process of identifying relevant similarities in moral issues.

In the domain of visual perception, for example, the difficulties encountered in reading for the first time require accommodation of the visual system to correctly discriminate the printed markings from peripheral phenomena. Difficulty in performing this preparatory task may well be symptomatic of a structural defect in one or both eyes e.g. astigmatism or presbyopia. Assuming the latter diagnosis, on that information alone, the possibility of reading would offer no affordance, so that any further attempts to read with such defective eyesight would be rationally incoherent in the light of that information. But if the defect can be corrected by surgery, or with the aid of contact lenses or glasses, then the availability of these tools constitute new affordances for accommodating the visual system to the task of reading. For this reason, having *information* of these adaptive solutions is itself an affordance which may be more or less accessible in different locations, where factors such as the financial cost of correcting the defect, the availability and content of reading material would also arise as affordances whose ecological variability would make them candidates for rational selection.

Even when all the material affordances for the reading activity are ecologically abundant, the correct identification of distinct letters and words and the interpretation of their syntactic and semantic relations is essentially a task requiring the resolution of competing ambiguities. Sentences such as "I saw her duck" or "He went to the bank" can only be rendered coherent by the reader actively *selecting* the appropriate meanings of 'duck' and 'bank' by a process of elimination and confirmation of alternatives in meeting certain contextual constraints.[385] Thus, the possibility that 'bank' here refers to a river would be maximally constrained by the preceding sentence stating that "he needed more dough". However, if the following sentence were to state that "there he found a boat", the prior determination of 'bank' would no longer be maximally constrained to one coherent meaning until the reader is once more able to eliminate or confirm one of the two contenders by further semantic clues. For example, if the narrative next described the character giving an exorbitant tip after paying a ferryman to row him across a river, that information would suffice to confirm the validity of interpreting the 'bank' as that of a river. While not entirely eliminating the validity of the original hypothesis that "he went to the bank" might yet refer to the character soon withdrawing some money, the knowledge that he is apparently *not now* in need of it, but *is* near the bank of a river are contextual circumstances which have the combined effect of lending strong explanatory coherence to the second, revised interpretation.

Such epistemic constraints in the coherent interpretation of information can be classified into several areas, each governed by a set of principles that evaluate data by their conformity to various 'discriminating constraints', as described in Thagard's comprehensive analysis.[386] He identifies five primary forms of coherence: explanatory, analogical, deductive, visual and conceptual. Although he

also adds deliberative coherence as an additional element applied to ethical and political decision making, its constraints are essentially the same as those determining the explanatory coherence of beliefs, but translated into teleological terms which are distinct enough to warrant their redescription in the domain of action.[387] In Thagard's system, the constraints of explanatory coherence are defined by the priority of acceptance given to the observational evidence for a hypothesis. This inductive principle thereby sets positive constraints upon the probability of alternative hypotheses, by the degree to which they are confirmed by the observations or by other hypotheses. Negative constraints are determined by contradiction and competition between propositions that cannot be linked to generate explanatory coherence.[388]

In the above task of resolving the ambiguity of competing word meanings, explanatory coherence is negatively constrained by the disconnection between the propositions that 'bank' refers to: (a) a riverbank *and* (b) a financial institution in the *same context*, while the positive constraints in support of accepting (a) are supplied by the priority given to the evidence of the river crossing. A further negative constraint in the selection of (b) is imposed by the contradictory relations between the references to: (a) the character's needing more 'dough' and (b) the generous tip he gives to the ferryman shortly after. While these two possibilities are not, of course, contradictory in deductive terms, their explanatory incoherence is itself derived from incoherence in the regular association of concepts. These associations may be objectively determined or constrained by physical laws and properties, as in the earlier example of the antagonism between fire and water, or they may be the expression of practical constraints imposed by regulative norms or conventions.

In this particular case, if the narrative in question appears to follow the stylistic conventions of a old-fashioned detective novel, it is the slang typically used in this genre that provides the maximal positive constraints for this conventional association of the concepts 'dough' and 'money', even where the contextual ambiguity remains strong, as would occur if a character enters a bakery shouting "Gimme all your dough!" If the baker then responds by handing over his baking dough which is duly accepted by the robber, this incongruity, which is itself a defining feature of humour, would here be incongruous in the context of a serious detective novel. On Thagard's account of humour "… incongruity is the incoherence between the initial coherent expectations in an utterance or situation and the final coherent interpretation after something surprising occurs".[389]

Thus, the incongruity of that particular scene would soon be resolved by its connection with other comic developments in the same vein, as would be the case if, during the aforementioned robbery, a hippy were to enter the shop and demand in a drawling voice: "Listen, man. I like need some bread, man! Like real bad!", whereupon the baker hands him a mouldy baguette. But if the surprising incongruity in the first scene is not repeated elsewhere in the narrative, or is not otherwise explicable as a dream episode or in terms of some surrealistic literary device, its presence in the novel would indeed be apprehended as *out of place* in the detective story genre. For this reason, this ecological incoherence also extends back in time to the process of editing, publication and even advertising, if the book has been wrongly categorised and marketed as a traditional detective novel, when it is in fact a pretentiously experimental 'stream-of-consciousness' detective-cum-fantasy hybrid, intentionally aimed at confounding the expectations of both reader and editor alike.

These conceptual conventions and normative expectations thus serve to identify coherent relations within specified ecological domains where different tasks generate different affordances. Having purchased the book with the hope of being presented with a mysterious crime to be unravelled, that hypothetical probability is nullified by the actual content of the material, so that its production offers negligible affordances for that reader's specific goal. But its lack of redeeming features in this particular respect thereby produces an instrumental void in the book's affordances which may now be used in the production of alternative ends. If this disappointed reader happened to be in the same position as the bushwalker searching for firewood, he may yet find genuine warmth afforded by the book in following Hume's advice on other tendentious tomes: "Commit it then to the flames"[390]

It is in the evaluation of these affordances that deliberative coherence has its function and the principles by which it operates are reminiscent of Dewey's approach to problem solving described earlier. However, Thagard's methodology is especially illuminating by providing a more precise analysis of the interplay among individual elements involved in the deliberative process, including the support which they derive from the other forms of coherence in his system. In introducing these different forms of coherence, it was noted that deliberative coherence is itself built upon the same epistemic principles that determine explanatory coherence, only reoriented to the task of accomplishing goals. Thus, while explanatory coherence is essentially manifested in: (a) the fitness of the relations among different hypotheses and (b) the priority assigned to the evidence for them, in deliberative coherence these are redefined to reflect constraints upon: (a) the fitness of the relations among different means and ends and (b) the priority assigned to certain intrinsic goals.[391]

In Thagard's terminology, deliberative coherence among means and ends is judged in terms of their mutual 'facilitation', which is precisely the same process described by Dewey's theory of valuation, wherein means and ends are not evaluated and selected independently of each other, but on the basis of their mutually constraining affordances.[392] In conceiving intrinsic goals as "ones that an agent has for basic biological or social reasons", the empirical justification for their priority can be seen as derived from that same priority in explanatory coherence and therefore subject to reevaluation and revision in the light of their observed effects.[393] Thagard's example of the need to restrain one's appetite for unhealthy foods in the interests of one's health certainly shows the incompatibility of two desires that are intrinsically *related*, but in characterising intrinsic goals as 'final ends' in distinction to instrumental goals, he commits the same error noted earlier in the discussion of basic goods in the theories of Aristotle, Hume and Mill. Good health here is rightly given priority over satisfying one's hunger indiscriminately because it facilitates further goals that would not be achieved by feasting on a plate of doughnuts.[394] However, in motivating the organism to eat, hunger is also only a *means* of facilitating that biological goal which cannot be considered final, as it is no less *instrumental* to the organism's continued pursuit of *further* activities.

Instead of equating intrinsic goals with the dubious notion of 'final' ends, Thagard's definition can be readily amended to reflect Gewirth's recognition of the basic goods that *essentially* facilitate the production and realization of all other goals. Thus conceived, freedom and well-being demonstrate their intrinsic priority as means, while still retaining their priority as immediate ends that must be secured in order to serve that instrumental function. In addition, this conception of their priority

reinforces the empirical connections between the principles of goal priority, facilitation and incompatibility in Thagard's analysis. Their priority not only facilitates all other goals but also provides a fundamental, substantive criterion for rejecting goals which are demonstrably incompatible with the preservation of those basic goods. Applied to Thagard's example, as long as the person has further goals which could *not be afforded* if all the doughnuts were consumed, then the decision to eat them would be rendered incoherent by its incompatibility with the level of well-being that is required to achieve those other goals.

Thagard's principles of goal priority, facilitation and incompatibility have already been highlighted in assessing the role of Gewirth's two conditions as grounds for his Principle of Generic Consistency. Like Thagard's, Regan's example of the aspiring ballet dancer who enjoys smoking raises all of the above factors in assessing rational coherence. But in addition to these central factors and their dependence on principles of explanatory and conceptual coherence, the process of rational deliberation also involves elements of deductive coherence in Thagard's account, though not constrained by the stringent procedures of formal logic as deduced in the form of *modus tollens* and *modus ponens*.[395] Nevertheless, like Gewirth and Hare, Thagard recognises the principle of non-contradiction as an elementary condition for the logical consistency of moral judgements. As an illustration, he cites the contradiction in asserting that "capital punishment is justified for heinous murders", whilst also maintaining that having committed such a murder does *not* warrant the execution of the person responsible in a particular case.[396]

While not explicitly invoking the principle of universalisation in support of the inconsistency here, Thagard's description of the relations of entailment between the two propositions clearly indicates his recognition that all cases of heinous murder are governed by the prescriptive principle. His agreement with Hare is further demonstrated by the distinction he draws between the qualification defined by "heinous murders" in contrast to the general claim that "capital punishment is wrong"[397] Here, the negative constraint between the general ethical principle and the exceptions it allows is not strictly deductive, as they can both be maintained without contradiction i.e. the exclusion of 'heinous murders' still permits the inclusion of many other classes of criminal acts. But while not invalidating the general claim, the acknowledgement of such exceptions is sufficient to generate deductive constraints in this less rigorous or 'soft' sense, as Thagard describes it.[398] By contrast, if the general principle were in fact intended as a universal claim, then by definition, it *cannot* admit any exceptions.

In keeping with the principle of universalisability, Thagard also notes the mode of reasoning by analogy to be important in the identification of the relevant similarities which establish coherent relations between inferences. Analogical coherence is essentially defined by the symmetrical correspondences that can be represented in the structural comparison of phenomena, be they physical or mental objects, events or processes. Thagard cites a particularly apt example in the analogous operations of natural and artificial selection in Darwin's theory.[399] As is often the case with scientific hypotheses, Darwin was, to some extent, led to postulate the workings of a law-like process in nature from his observations of a similar process directed by human intervention i.e. the techniques of selective breeding in animal husbandry and horticulture.[400] Interestingly, Marx remarked that Darwin's theory was born of an inevitable analogy founded upon his observations of the ruthless effects of

competition promoted in English society at the time.[401] Although intended as a derisive criticism of Darwin's theory for purporting to describe as a law of nature what Marx regarded as only a utilitarian law of human artifice, this objection has since proven to be another example where the order of inference has been misconstrued. Darwin's theoretical musings may well have been influenced by his awareness of the effects of unfettered capitalism in fostering a ruthless competition for resources. However, in the previous century, Adam Smith had already drawn attention to the autonomous operation of the laws of supply and demand in processing the aggregate value of individual transactions, as if by an 'invisible hand'.[402] To the extent that such basic economic laws evidently operate in the same causal manner, *independently* of the content in the items exchanged, they would be *expected* to have similar effects in any domain where resources are sufficiently scarce to require competition. Therefore, to hypothesise the operation of analogous economic processes in nature cannot be characterised as an anthropomorphic attempt to read such laws into the natural world. On the contrary, as opposed to the pseudo-scientific absurdity of Marx's theory of dialectical materialism in nature,[403] Darwin's 'reading' of natural selection has proven to be an accurate reading *of* a natural process whose artificial analogues can be read as *evidence* of its successful agency in nature.

As Thagard notes, the similarities drawn by analogy are ordinarily constrained by their 'mapping' in visual or semantic terms.[404] Although the meanings of concepts are often quite vague when considered in isolation, this lends them a greater analogical facility for mapping a wide variety of semantic correspondences in many areas. For this reason, it cannot be objected that Darwin's use of the term 'selection' is only a metaphorical description that is inappropriate or misleading when applied to a mechanical process devoid of intentionality. As long as the term is used to locate close correspondences in the *conditions* of its operation, so that it produces predictably similar effects in natural or artificial conditions, intentionality is not necessary to establishing those correspondences. Evolution by natural selection occurs via the combined operation of three conditions that are both necessary and sufficient: reproduction, heritability and variation.[405] Therefore, selection processes are bound to occur *wherever* those conditions hold, regardless of the source of the conditions, the mode of transmission, or the objects of selection. These same considerations therefore apply equally to the semantic interpretation of the conditions themselves. For example, as 'reproduction' need not be sexual for evolution to occur, the term 'replication' may be used elsewhere instead.[406] Likewise, beliefs, values and norms may be inherited through the media of language and culture and their variation conditioned by the social environment. But as the conditions of heritability, replication and variation are still present, the combined effects of their operation is analogous to that which occurs in nature.

As the concept of selection shows, the common effects observed in analogical coherence are significant for their explanatory power which Thagard therefore cites as a further constraint upon the practical purposes to which the analogy may be coherently applied.[407] Applied to ethical considerations, this principle provides a powerful tool for discriminating among meta-ethical theories, while also imposing constraints upon the explanatory coherence of normative claims and specific moral judgements. The so-called 'Boo! Hurrah!' emotivist theory[408] that moral judgements are analogous to expressions of emotions or attitudes, devoid of any cognitive control or critical function, is one example vulnerable on these grounds, as the analogy fails to yield any correspondences that might

suffice to explain the role of moral principles, conscience, or rational argumentation which frequently work to *inhibit* the free expression of emotions .

The deliberative coherence of normative claims is also restricted by the explanatory value of the semantic or visual correspondences that are purported to hold between judgements. In the case of capital punishment, Thagard shows how certain correspondences can be construed to generate constraints that are analogically consistent with the claim that "capital punishment is wrong". The belief that capital punishment is analogous to abduction and murder can be readily constructed by identifying the relevant semantic similarities in the terms 'kill' and 'execute', while the abductor's coercive restraint of the victim generates positive syntactic correspondences from 'abductor' to 'State' and 'victim' to 'prisoner'.[409] The coherence of this analogy does not, of course, prevent alternative correspondences being established in support of coherence in opposing views of the same issue. Thus, Thagard notes that an analogically coherent argument in favour of capital punishment can be advanced by locating relevant similarities in the rationale for killing in self-defence and execution as a form of pre-emptive self-defence.[410] Alternatively, the practice may be defended on the basis of a *prima facie* principle of desert or retributive justice. However, as the rationale for *any* principle imposes coherence constraints on the validity of the beliefs which may be cited in its *support*, it cannot be justified independently of those supporting beliefs. If, for example, the person defending capital punishment on the basis of desert should also hold the belief that "all human life is sacred", then there is an obvious deductive incoherence in that person's reasoning.

Although competing analogues can always be identified and constructed to lend some internal coherence to alternative viewpoints on many ethical and political issues, that coherence will be further constrained by its conformity with the deductive and explanatory elements that also contribute to the deliberative coherence of a specific judgement or principle. The question of the morality of capital punishment is also answerable to the empirical evidence that constrains the explanatory coherence of the causes and effects cited as premises for advocating certain prescriptions. For instance, if it is argued that capital punishment is justified on the grounds that it acts to deter the heinous crimes that would otherwise be committed if the punishment were any less drastic, these are predictions which can be scientifically tested to determine their efficacy.[411] But as this experimentation can only be realistically conducted by actually having the punishment publicly instituted and witnessed, that decision would thereby promote negative constraints in its possible coherence with the priority given to the intrinsic biological and social goals of those subjected to such trials.

As deliberative coherence is primarily concerned with determining the facilitation of actions towards goals, incompatibilities between goals are resolved by that very principle. By definition, in ethical deliberations the incompatibilities necessarily include the goals of all those affected by the implementation of any prescription or policy, even if it is only temporarily enacted in order to determine its explanatory value for some other goal that could then be maximally facilitated. Therefore, even if the experimental introduction of capital punishment in a particular State might indeed prove to be a powerful deterrent, the considerable explanatory coherence that can then be employed to maximal effect in deterring heinous crimes is nevertheless likely to yield substantial deliberative *incoherence*, by ignoring the likely impact of the decision on other concerned parties.

In addition to the empirical question of establishing the guilt of the prisoner "beyond reasonable doubt", the authorisation of a single act of execution is liable to produce a myriad of changes in the psychological and material conditions afforded to many others who bear no responsibility for the particular crime. These are also questions which are amenable to investigation, by enquiring into the individual circumstances of the parties affected and through psychological and sociological research into the likely effects upon families, friends and associates of both the person condemned and the victim of the crime.[412] Even if reliable scientific research on the probable causes and effects of a proposed moral or political prescription is inconclusive, or cannot be conducted without concern for those affected, information derived from analogous events or processes may be cited to lend sufficient coherence to a moral argument on the issue. In the case of capital punishment, the psychological effects upon different parties can be reasonably inferred from the vast number of historical precedents of similar practices found in many cultures. Insofar as relevantly similar moral concepts and normative values are invoked to defend the practice and the societies that are the source of the analogy also exhibit a corresponding freedom of individual expression on such questions, such information can then serve as hypotheses whose analogical and explanatory coherence can then be evaluated against the intrinsic priority given to the facilitation of biological and social goals in deliberative coherence.

The more positive constraints that can be forged between elements across each relevant type of coherence and the fewer the negative constraints, the greater will be the overall coherence displayed by a particular moral judgement, as the outcome of this deliberative process.[413] Therefore, on the basis of the aforementioned factors relevant to the morality of capital punishment, imagine that actual or analogical evidence were to show: (i) moderate support for the deterrence hypothesis, (ii) quite significant evidence for serious, long-term losses in psychological and socio-economic well-being borne by families of those executed and (iii) slight increase in psychological benefits afforded to some crime victims by the experience of retribution. While (i) and (iii) can be coherently maintained, they cannot forge any deductive, explanatory or analogical links to generate positive constraints in their favour. By contrast, in addition to the condemned person's own deprivation of life, the substantial multiple losses of intrinsic capacities in (ii) generate strong negative constraints against the deductive and deliberative coherence of prescribing capital punishment to produce the hypothetical benefits in (iii) for which explanatory coherence is weak.

A more thorough account of relevant ethical considerations on this issue would also require analysis of various proximate and longitudinal causal factors that can be hypothesised as influencing the rational deliberations of all the parties concerned. Ethical deliberations in judicial proceedings are especially instructive in exposing the psychological factors that predispose contesting parties to interpret information to justify incompatible perspectives, while also furnishing some principles for setting formal constraints upon the prejudicial conduct and content of the enquiry which may therefore serve to guide some analogical deliberations on social and political ideals under a social contract. But this objective can only be approached by first highlighting the conditions required for establishing a level of reflective equilibrium within the systematic moral judgements of citizens, one that is best fitted to informing and sustaining their commitment to that long-term social task. Moreover, these

multicoherence standards may also function as a critical tool for the identification and adaptive reconstruction of moral and political rules and institutions, so that the *external* equilibrium they express continually serves to promote the *internal* reflective equilibrium achieved by the critical reflections of autonomous agents.

(iii)   Reflective Equilibrium:

Thagard's systematic approach is well designed for both detecting incoherence and promoting maximal coherence across a wide spectrum of beliefs, values and ideals in a manner that is nevertheless fully consistent with logical and epistemic principles that have evolved to constrain the interpretation of information in terms of local rather than global maxima.[414] Therefore, while his "multicoherence" theory certainly provides a systematic global account of the maximal scope for establishing a 'wide' reflective equilibrium over all the elements within a singular moral perspective, Thagard concedes that the normal process of moral reasoning is much more economical and *ad hoc* in attempting to fit available information into a subset of local coherence constraints.[415] As noted earlier, many 'fast and frugal' heuristics are perfectly coherent with the task of evaluating information that affords an immediate risk assessment and response to local problems and is often *superior* to judgements whose wide coherence is dependent on a time-consuming, global analysis of all available information.

Nevertheless, while such heuristics and biases succeed by sharply narrowing the focus of expectations and probabilities to the most proximately defined data, their superiority in these myopic tasks prevents them from detecting and evaluating additional information afforded by a wider focus of concern. Of course, like the smoker who is intent upon ignoring the significance of global data even when it threatens his own life, reflective equilibrium is liable to remain intentionally narrow, so as to avoid acknowledging information that challenges the coherence of existing beliefs and habits. As long as the smoker entertains background beliefs that admit the validity of evidence for the harmful risks of smoking, he cannot coherently maintain his 'considered' belief that his habit is not harmful, when that belief is only founded upon an *ad hoc* hypothesis drawn from knowledge of a few local characters whose health and longevity conflicts with the global evidence. Given the serious effects of ignoring vital information in ethical and political questions, moral reflection must be aimed at sustaining a wide equilibrium where "background beliefs" are given as much consideration as those "considered moral judgements" whose coherence would otherwise be only narrowly constrained to accord with some vague moral theory which can easily be reconstructed to simply endorse those judgements without revision.[416]

One of the fundamental objections to Rawls' defence of reflective equilibrium is its reliance on prevailing moral 'intuitions', which are accorded a privileged status as judgements assumed to be 'considered' enough to guide the possible coherence of new, alternative beliefs or values. Singer, in particular, justly decries the naïvety in giving credence to notions that may be nothing more than prescientific superstitions or outmoded dogma which only survive by *avoiding* severe scrutiny.[417] Historically, the most abhorrent practices and prejudices have been able to survive scrutiny by masquerading as self-evident moral 'truths' that no right-minded citizen could possibly question. In

considering the rational scope of moral reflection, Hare notes approvingly of Brandt's use of the term 'rational' "to refer to actions, desires, or moral systems which survive maximal criticism by facts and logic."[418] By drawing attention to the role of facts as epistemic constraints in the coherence of moral systems, rather than as descriptions of moral truth, this level of critical reflection is able to expose the 'error' that is often responsible for the privileged position assigned to commonly shared moral intuitions. Accordingly, only those intuitions which *continually survive* this process will qualify as genuinely *considered* moral judgements.

Although some long cherished moral beliefs are liable to be rendered incoherent by their dependence on falsified empirical hypotheses, this need hardly lead to the kind of nihilistic moral vacuum that is often invoked in defending a conservative approach to reflective equilibrium. After long and thorough scrutiny, many other intuitively acceptable beliefs are likely to retain their moral priority, for precisely those reasons identified by Thagard in reference to the primacy of common biological and social goals. In adopting this critical approach to the assessment of moral beliefs and values, it may well be argued that moral concepts themselves may yet be dispensable, in favour of much greater objectivity in determining the conditions which would suffice to warrant the provisional adoption of a given social norm or ideal. In fact, this is what Dewey himself suggests in the following remarks: [419]

> Serious social troubles tend to be interpreted in *moral* terms. That the situations themselves are profoundly moral in their causes and consequences, in the genuine sense of moral, need not be denied. But conversion of the situations investigated into definite problems, that can be intelligently dealt with, demands objective *intellectual* formulation of conditions; and such a formulation demands in turn complete abstraction from the qualities of sin and righteousness, of vicious and virtuous motives, that are so readily attributed to individuals, groups, classes, nations…And such formulation is the sole mode of approach through which plans of remedial procedure can be projected in objective terms. Approach to human problems in terms of moral blame and moral approbation, of wickedness and righteousness, is probably the single greatest obstacle now existing to development of competent methods in the field of social subject matter.

Even the general, substantive value in satisfying such common desires as hunger must still be subjected to critical reflection in order to assess its possible value under specific conditions, as in the case of being presented with a plate of doughnuts. Furthermore, as moral reflection specifies conditions which require the assessment of those same expressions of hunger in others, then whatever value is found in the conditions which permit the satisfaction of one's own hunger must also be granted to those in the same conditions.

Imagine this same person indulging in a plate of doughnuts while watching television. He intuitively endorses the general moral principle that "no-one should be allowed to go hungry" and when his viewing is interrupted by an advertisement requesting donations for starving children in some desolate and distant land, the images temporarily induce him to cease munching. But as the images dissolve, his conscience is soon silenced by the perceptual proximity of the remaining doughnuts, which succeeds in restoring reflective equilibrium by reducing the focus of attention from a wide to narrow specificity, so that the satisfaction of his own hunger is presently interpreted as one instantiation

of the general moral principle he asserts. Thus, he reaffirms his commitment to the intuitive moral belief that "no-one should be allowed to go hungry" by dutifully eating his fill of doughnuts, an act which is given further coherence by its agreement with his equally considered endorsement of the prescription expressed by the slogan "waste not want not". While this facile, *ad hoc* reflection is capable of sustaining a narrow equilibrium to allow continuing engagement with current goals, the coherence it obtains is destined to be short-lived, as it succeeds only by constantly deferring consideration of conflicts that may arise to threaten the coherence of such beliefs in their wider implications. In this example, narrow equilibrium effectively *suppresses* the person's moral conscience by subverting the general focus of his principles to give priority to his present concerns, as a *beneficiary* of those principles. But unless his conscience is sufficiently robust to soon remind him of the contradiction entailed by failing to act upon that same principle, with the urgency demanded by those whose lives are threatened by such neglect, narrow reflective equilibrium is liable to remain captive to an equally narrow self-interest.

> Every Man's Conscience Is Vile And Depraved/You Cannot Depend On It To Be Your Guide/When It's You Who Must Keep It Satisfied
>
> (Dylan - *Man In The Long Black Coat*)[420]

As these lyrics succinctly attest (aside from the poetic licence evident in the excessive cynicism of the first line), the strongest moral intuitions may be prone to corruption by personal desires, to such an extent that one's own conscience not only comes to tolerate frequent violations of principles that it should serve to uphold, but may positively and consistently recommend the pursuit of self-interested goals as a *duty*. This is reminiscent of the criticism that Bentham's promotion of the pleasure principle is "a doctrine worthy only of swine",[421] which is also forewarned in Aristotle's recognition of the moral dangers in being guided exclusively by pleasurable ends.[422] This suspicion of the inherent malleability of the private conscience substantially undermines the surety required of 'considered' judgements in their capacity to function as intuitively reliable standards for evaluating the acceptability of other prospective moral principles, prescriptions or theories.

Given that even the most intuitively acceptable moral principles cannot be relied upon to provoke sustained reflection to assess their tenability in a widely constructed moral system, a self-interested commitment to such principles will inevitably favour a method of reflection strategically designed to reinforce that commitment. This *conservative* approach to maintaining equilibrium makes a mockery of the role and extent of critical reflection in moral questions, reducing the deliberative task to one that is literally *hypocritical*. Therefore, nothing less than a radical conception of reflective equilibrium will suffice to allow for the effective revision of long-standing principles, where these are increasingly found to be poorly adapted to the tasks under their direction.[423] In the above case, if repeated exposure to information detailing the continuing hunger of many thousands of people still fails to prompt action consistent with the professed belief that 'no-one should be allowed to go hungry', then unless there are practical obstacles which might temporarily hinder attempts to contribute to that effort, that belief no longer sustains a coherent function in the direction of moral deliberation or action.

Although weakness of will often plays a role in undermining genuine attempts to adhere to one's own moral convictions, such failure may also signal a weakness in the support such convictions receive from other values or ideals in the same moral system. In this example, the swift appeal to the prudential 'waste not want not' principle facilitates an indefinite suspension of any serious attempt to reflect upon the strength of concern for the hunger experienced by others, thereby serving only to avoid exposing the latter belief to a system of wide reflective equilibrium which would reveal its tenuous, if not contradictory relations with the majority of other concurrent beliefs and principles. If widely critical self-examination of those beliefs and principles showed a remarkable consistency among them, while also conforming to the tenets of an inchoate theory of ethical egoism, then it is the intuitive moral concern for other's hunger which is incoherent in such company. As such, instead of functioning to promote considered judgements of the interests of others, that isolated intuition barely survives as the 'weakest link' in an inhospitable moral system.

This example highlights the difficulties in relying on the method of reflective equilibrium to accurately identify those elements of the moral system which might best contribute to maximising the rational coherence expressed by any particular judgement. However, as most of these interpretive problems arise as a result of affirming the intrinsic value of existing beliefs and ideals without due investigation of their tenability as guiding hypotheses, they need not pose a threat to the task of fully testing their coherence against all elements of a system designed to establish wide equilibrium.

---

[338] Gewirth, *Reason and Morality*, pp.104-105

[339] James, William. *Selected Writings* ed. G. H. Bird (London & Vermont, Everyman, 1995), p.353

[340] Ibid., p.345

[341] Philip G. Zimbardo, *Psychology and Life*, Thirteenth ed. (New York, HarperCollins, 1992), p.603

[342] Ibid.

[343] Kierkegaard, op. cit., Vol. 2, p.218

[344] Ibid., p.244

[345] Haworth, op. cit., p.55

[346] R. M. Hare, *Moral Thinking: Its Levels, Method, and Point* (Oxford, Clarendon Press, 1981), p.172

[347] Ibid., pp.171 & 181

[348] Though not attributable to Buridan himself – See Michael Bratman, *Intention, Plans, and Practical Reason* (Cambridge, Mass., Harvard University Press, 1987), p.11 note 23)

[349] Ibid., p.11

[350] R. M. Hare, *Freedom and Reason*, pp.87-88

[351] John Dewey, *Human Nature and Conduct* (New York, Henry Holt, 1922), pp.10-12, 243 & 296

[352] Hare, *Freedom and Reason*, pp.38-40

[353] Ibid., p.40

[354] Ibid., p.15; *Language of Morals*, pp.111-115

[355] Hare, "What Makes Choices Rational?" in *Essays in Ethical Theory* (Oxford, Clarendon Press, 1989), p.37

[356] Ibid., pp.36-37.

[357] Hare, *Freedom and Reason*, pp.153-154

[358] Mackie, *Ethics: Inventing Right and Wrong*, pp.87-88

[359] Hare, *Freedom and Reason*, p.32

[360] Mackie, *Ethics: Inventing Right and Wrong*, p.42

[361] J. L. Mackie, "The Law of the Jungle: Moral Alternatives and Principles of Evolution," in *Issues in Evolutionary Ethics*, pp.165-177

[362] e.g. in discussing the merits of Gewirth's argument for the PGC, Bernard Williams agrees that "if in my case rational agency alone is the ground of a right to noninterference, then it must be so in the case of other people", affirming that "it rests upon the weakest and least contestable version of a "principle of universalizability" " - *Ethics and the Limits of Philosophy*, (London, Fontana, 1985), p.60

[363] Mackie, *Ethics: Inventing Right and Wrong*, p.83

[364] Ibid., pp.90-97

[365] Peter Singer, "Reasoning towards Utilitarianism," in *Hare and Critics: Essays on Moral Thinking*, ed. D. Seanor, N. Fotion, et al. (Oxford, Clarendon Press, 1988), p.150

[366] R. M. Hare, *Sorting Out Ethics* (Oxford, Clarendon Press, 1997), p.99

[367] R. M. Hare, "Universalisability," in *Essays on the Moral Concepts* (Berkeley & Los Angeles, CA; University of California Press, 1972), pp.16-17

[368] John Dewey, "Valuation and Experimental Knowledge," in *The Essential Dewey*, vol. 2, p.275

[369] Dewey, *Human Nature and Conduct*, p.208

[370] John Dewey, "Valuation and Experimental Knowledge," in *The Essential Dewey*, vol. 2, p.284

[371] John Dewey, "Philosophies of Freedom," (1928) in *The Essential Dewey*, vol. 2, p.304

[372] Hare, *Freedom and Reason*, pp.90-91

[373] Ibid., p.92

[374] Ibid., pp.93-94

[375] Hare, *Moral Thinking*, p.46

[376] Hare, *Freedom and Reason*, p.106

[377] Hume's famous remark that: "Reason is, and ought only to be the slave of the passions…"(*A Treatise of Human Nature*, II. III. III. p.415) may well be ironically cited as an egregious breach of the is/ought distinction of which Hume is, of course, cited as originally proposing in a latter section of the same work (i.e. III. I. I. p.469). But the usage of 'ought' here may be more coherently interpreted to mean 'could'.

[378] Gigerenzer, op. cit., pp.241-242

[379] Ibid., pp.228-230

[380] Ibid., p.263

[381] Skyrms, op. cit., p.85

[382] Skyrms, op. cit., pp.98-100

[383] Edward S. Reed, *Encountering the World: Toward an Ecological Psychology* (New York, Oxford University Press, 1996), pp.18-19

[384] Edward S. Reed, "The Intention to Use a Specific Affordance: A Conceptual Framework for Psychology," in *Development in Context: Acting and Thinking in Specific Environments*, ed. R. H. Wozniak & K. W. Fischer (Hillsdale, N.J., L. Erlbaum, 1993), pp.45-57

[385] Paul Thagard, *Coherence in Thought and Action* (Cambridge, Mass. ; London, MIT Press, 2000), p.81

[386] Ibid., p.63

[387] Ibid., pp.128-129

[388] Ibid., pp.43 & 63; Note Thagard's use of the term 'constraint' to also refer to +coherence in denoting connections in favour of one over other alternatives, in the sense that these +connections *constrain* the coherence of alternatives.

[389] Ibid., p.206

[390] Hume, *Enquiries Concerning Human Understanding,* XII. III

[391] Thagard, op. cit., pp.43 &128

[392] Dewey, *Theory of Valuation*, pp.29-33

[393] Thagard, op. cit., p.127

[394] Ibid., p.129

[395] Ibid., p.132

[396] Ibid., p.133

[397] Ibid., pp.132-3

[398] Ibid., p.132

[399] Ibid., pp.49-50

[400] Darwin, Charles, *The Origin of Species*, First ed. (1859), Ch.1.

[401] cited in Skyrms, op. cit., p.45

[402] Adam Smith, *The Theory of Moral Sentiments*, vol. 1, (Oxford, Clarendon Press. 1976), IV. ii. 9, p.456. See also Smith's *An Inquiry into the Nature and Causes of the Wealth of Nations* (Oxford, Clarendon Press, 1976), IV. i. 1, 10

[403] Karl Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (London, Routledge & Kegan Paul, 1974), pp.331-335

[404] Thagard, op. cit., p.50

[405] See Elliott Sober & David Sloan Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior*, (Cambridge, Mass., Harvard University Press, 1998), pp.104-105

[406] George C. Williams, *Natural Selection*, p.12

[407] Thagard, pp.50-51

[408] See Charles L. Stevenson, *Ethics and Language* (New Haven, Yale University Press, 1944) as most representative of emotivism

[409] Ibid., pp.138-139

[410] Ibid., p.139

[411] Ibid., pp.129-130

[412] Ibid., pp.130-131

[413] Ibid., p.126

[414] Ibid., p.141

[415] Ibid., pp.141-2

[416] Michael R. DePaul, *Balance and Refinement: Beyond Coherence Methods of Moral Inquiry* (London, Routledge, 1993) pp.20-21

[417] Ibid., p.26

[418] Hare, *Moral Thinking*, p.214

[419] John Dewey, "Logic: The Theory of Inquiry," extract in J. Gouinlock (ed.) *The Moral Writings of John Dewey* (New York, Macmillan, 1976), pp.170-171; see also *Human Nature and Conduct*, Part IV, Section IV; *The Quest for Certainty: A Study of the Relation of Knowledge and Action* (New York, Capricorn Books, 1960), pp.271ff. In this context Dewey also observes that the "framework of moral conceptions" is a permanent fixture of the sociality of human nature, so his concern is rather with "particular aspects of morals…that are often, in their actual manifestation, defective and perverted." – Dewey, "The Moral Self," in *The Essential Dewey*, Vol. 2, p.354

[420] From the album *Oh Mercy* ©1989

[421] Mill, op. cit., p.7

[422] Aristotle, *The Nicomachean Ethics*, II. ix. 6

[423] DePaul, op. cit., pp.38-39

CHAPTER SEVEN

THE PROBLEM OF SPECIFICITY:
AN ECOLOGICAL AND DYNAMIC RESPONSE

Through their combined operation, the foregoing principles and methods of rational coherence impose restrictions upon the tenability of particular judgements within an individual moral system. Nevertheless, in spite of all such constraints, a system may attain maximal internal coherence encompassing a wide spectrum of beliefs, values and principles, yet fail to be similarly constrained by its interpretation of the constituents of *external* phenomena. Although a set of moral beliefs may be mutually consistent and perfectly compatible with values and principles that are equally effective in practice, unless some of those beliefs can be relied upon to accurately locate objectively meaningful correspondences in the environment, the entire system is bound to admit significant and regular errors in judgement. Thus, the problem of determining the continuing tenability of considered moral judgements is compounded by the inherent *indeterminacy* of objective meanings and values that is permitted by a system which eschews foundationalist attempts to guarantee certainty of interpretation. Insofar as moral propositions do not assert objective facts but only identify hypothetical possibilities for action, the epistemic systems which inform them are, of course, designed to operate with a degree of latitude that allows for substantial revision of personal values. As such, instead of guaranteeing certainty, moral systems guarantee the opportunity to redefine and reconstruct beliefs to reflect a

corresponding change in values. Therefore, in contrast to a coherence theory of truth, the epistemic basis for the coherence of moral beliefs is open to frequent and radical reinterpretations of objective means and ends, as revealed in the light of newly discovered values or ideals elsewhere in the system.

However, the interpretational latitude of any moral system must still be constrained in its capacity to correctly and consistently locate and evaluate objective properties of phenomena, by the extent to which they afford or obstruct the realization of chosen values. For this reason, relations of purely internal coherence can tolerate wholesale errors of judgement that would soon be eliminated if tested against epistemic principles where a special priority is given to reliable *observations* in fixing the determinate content of beliefs. But this is precisely where Thagard's 'multicoherence' theory goes beyond the internal boundaries of pure coherence, by recognising the primacy of empirical evidence in providing initially plausible grounds for a reliably objective representation of phenomena. In contrast to foundationalist attempts to build a system upon a set of self-justifying *a priori* beliefs or principles which are regarded as indubitable and immutable, the priority of experimental evidence in Thagard's account of explanatory coherence does not function as an absolute constraint to which all other elements must conform, as such evidence is itself constrained by *its* coherence with whatever analogical, deductive, visual and conceptual elements also contribute to the validity of any hypothesis.[424] Therefore, while its empirical approach avoids the 'no contact with reality' objection which besets pure coherence theories, Thagard's system, like Bonjour's earlier proposal, also avoids the 'no credibility' objection which besets foundationalist claims to establish the absolute certainty of some beliefs, independently of their coherence with the rest of the system.[425] In this respect, Thagard acknowledges his theory as an elaboration of Haack's 'foundherentist' epistemology whose principles are best illustrated in the task of solving crossword puzzles.[426]

The reasoning employed to solve such puzzles does not proceed by focusing upon an isolated sequence of letters in one dimension (Across *or* Down) whose semantic construction is independent of meaningful sequences in the other dimension. Instead, each particular clue must be interpreted in the presence of other clues in both dimensions, which combine to fix the formation of words by their precise location. As more and more words are successfully combined to complete a section of the puzzle, their selection gains more and more credibility by the degree to which they contribute to the solution of the *entire* puzzle. Thus, their increasing validity is not founded upon any pre-established certainty that is unaffected by the information revealed in the rest of the puzzle. On the contrary, the valid selection of those first words gains warrant only through their coherence with the remaining words, which likewise are warranted by their coherence with all the other words in the puzzle at its current stage of completion. As a result, small clusters of letters and words will soon accrue a substantial number of confirmatory links that justifies their use in guiding the selection of other letters and words.

In contrast to the purely foundational approach where inferential justification proceeds in a one-directional linear fashion from an isolated set of axiomatic beliefs *a b c* to all other beliefs in the system, the foundherentist model proceeds in a dialectically organic manner where all beliefs in the system more or less actively contribute to the selection of the entire set, in accordance with the coherence principles described earlier. As Thagard notes, the success of computational models of

parallel information processing provides strong support for an holistic description of inferential reasoning that is "nonmystical, computationally effective, and psychologically and neurologically plausible".[427] However, unlike a computer program such inferential holism should not be construed as a complete or thorough system, where every belief is permanently retained or reactivated in response to one or other beliefs elsewhere in the system. Owing to limitations in the capacity to reliably store and retrieve information in the human brain, beliefs must be compressed into shared compartments so that only those with the most frequent and important application can be fully accommodated in long-term memory.[428] As a result, beliefs that are occasioned by objects and events perceived as insignificant or ephemeral, such as the chirp of a bird or the hum of traffic, will rapidly fade from memory to be replaced by those that are more salient to the task at hand.[429]

The attention to parallel sources of information in guiding decisions must therefore be limited to calling upon some cluster of beliefs, whose strong mutual coherence is given maximal credibility by their reliability in detecting only those meaningful aspects which are *specifically* salient to concurrent activities. As long as those aspects are objective properties of the actual environment in which human beings evolved, they are bound to be commonly recognised as such. As Quine neatly puts it: "Creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind."[430] As natural selection has attuned our cognitive capacities to detect a range of affordances and hazards that are common to most habitable environments, the objective value of many basic human interests is similarly attuned. For this reason alone, there is bound to be a range of objective values that are vital in their contribution to daily life in all surviving cultures. Accordingly, those heuristics and norms which are especially adapted to the maintenance of each individual's life must be given substantial credence in their capacity to induce responses that are at least *sufficient* to the completion of common biological goals. Moreover, as these rules evolved to guide activities within human societies, they are *specifically* designed to enable individuals to co-ordinate their efforts in locating and processing information that can be shared to mutual advantage.

With respect to the rational recognition of morality, Gewirth is willing to concede that: "Such evolution may well provide the necessary material background of our recognition, in that, if humans had not evolved as they have, with natural selection operating to foster the survival of certain human abilities, humans would not be able to make the moral judgments they do (sometimes) make on the basis of rational criteria." However, while granting that the evolutionary process may therefore be a necessary condition for moral behaviour, Gewirth adds that, in relation to ethics, this process alone "is not *specific* enough to account for its specific subject-matter, which involves the *differentiation* of the morally right from the morally wrong in human action and judgment."[431] Inasmuch as self-interested behaviour which runs counter to ethical concerns has also survived the selection process, it would indeed be foolish to expect such a mechanistic and *cosmic* process to favour the development of *exclusively* altruistic inclinations, especially where these are liable to promote the realization of ends that limit the altruist's own reproductive success i.e. by benefiting selfish mutants or invaders. Thomas Huxley, the greatest intellectual defender of Darwin's theory at the time of its publication, was similarly concerned to deflate its ethical implications, emphasising that the self-restraint induced by the ethical process is fundamentally opposed to the self-assertion that is rewarded by the cosmic process of

evolution in nature.[432] Huxley attributes ethical development to a gradual process of artificial selection, analogous to that produced by horticultural techniques used to cultivate a garden in the midst of a natural wilderness.[433] In this sense, ethical principles and methods are comparable to the tools by which humans are themselves able to tame their natural tendency toward untempered growth and proliferation, in the struggle to obtain sufficient space and nourishment to survive encroachments from the state of nature.

Yet Huxley also admits that this artificial process is itself made possible by the operation of the cosmic process, which encompasses and sustains the human intervention by providing and replenishing the raw material for the progressive cultivation of the 'moral sense' or conscience.[434] Similarly, notwithstanding his objection over the 'problem of specificity', Gewirth elsewhere acknowledges that: "To have the power of reasoned choice, of voluntary control over one's actions, could have been singled out by the evolutionary process as being more adaptive for individual and group survival than having only simple automatic reflex responses.[435] But as these powers can also be used in the service of maladaptive, socially harmful or self-destructive ends, Gewirth still insists that while necessary, they are not sufficient for the singular development of ethical action which requires the presence of further cultural influences and artifacts such as "language, property, education, wealth and income, self-respect, honor, dignity, and so forth."[436]

In Gewirth's own theory, this specificity in the content of moral value is directly and rationally determined by the degree to which these elements contribute to the freedom and well-being needed to maintain personal agency. But these cultural, material and institutional conditions, like the freedom and well-being they sustain, can *also* be directed towards *immoral* ends, despite their value in achieving specifically moral goals. Clearly, a person's material assets can be used to support specifically nefarious activities, but even those virtues which are implicitly defined by their normative moral usage can be deployed as vices for anti-social or Machiavellian ends. It should not be surprising that the moral value of self-respect, honour, dignity etc. is *not sufficiently* determined by the *generic* freedom and well-being they permit, as they describe personal characteristics which presuppose a degree of *self-restraint* that is incompatible with that generic conception. This was, of course, the essence of Regan's complaint against Gewirth's unjustified assumption that rational agents must value their prospective freedom and well-being, *regardless* of their specific purposes. Accordingly, it is Gewirth's *all-purpose* conception of freedom and well-being that lacks specific moral content in proceeding from the *unspecified* content of the agent's belief that '*E* is good'.

As 'honour' may be 'among thieves', a person may be motivated by an inflated sense of self-respect which has the effect of diminishing the dignity that might otherwise be assured by their ethical behaviour towards those in the wider society. As such, it is the *social* context and *normative* function of such values that defines their moral status. Therefore, the substantive goods which are essential for *moral* agency can only be discovered by enquiring into the ecological conditions which are both necessary and sufficient to produce and maintain these goods. Furthermore, insofar as the goods for prospective agency must be continually *reproduced* by agents themselves, the cultivation of normative moral virtues, principles and rules of decision-making must be guided by knowledge of the dynamic process of their mutual selection. When the evolution of prospective moral agency is examined in both

its dynamic and ecological dimensions, the indeterminacy of moral judgement and its content is much less problematic.

Following Reed's theory of ecological psychology, human thought typically evolves within a social "field of promoted action", wherein environmental regularities are observed, selected and internalised by individual agents for their value as normative affordances that are already adapted for a specific range of human ends.[437] These "specializations of time, place, events, objects, tools and so on"[438] are comparable to the 'ready-to-hand' meanings and values encountered in Heidegger's phenomenology. However, in contrast to their depiction as purely subjective and inauthentic representations of the scope of human thought and action, Reed's account gives full recognition to the *objective* value of cultural norms in facilitating the task of detecting stable and reliable patterns in the representation and use of information that is salient to a particular end-in-view. To be sure, most social groups and cultures are likely to promote some traditional codes of representation and action which have little or no pragmatic value, serving purposes which no longer reflect the wider range of prospective affordances available in the present environment. As Philippa Foot argues, many of the prescribed rules of etiquette exhibit this ossified structure[439] which would be totally superfluous, if they did not express at least some vestigial symbolism as a sign of the practitioner's co-operative disposition. On the other hand, many surviving customs do show a pragmatic purpose in specifying the means for ensuring that shared activities are guided by the *same* information. For instance, if the port is *invariably* passed to the left, everyone can expect to receive *equal opportunity* to partake, which would otherwise be a matter of chance distribution.

In their modest effects, such simple norms promote the prospective awareness and autonomy of agents in a specific field of action, by generating a common understanding in the interpretation of information relevant to success in that field. Therefore, in the task of selecting information appropriate to guiding behaviour, success depends upon agents developing the capacity to locate what Bogdan describes as "the right norms or scripts or departures from them in the cultural space."[440] In the examples already discussed, the different environs of bushland, supermarket, or smoky pub afford a range of opportunities for human action that is more or less structured to accommodate agents' normative interests. A trail in the bush, an aisle in the supermarket, a seat on a train, a cigarette dispenser in a pub – these artifacts describe objective affordances which facilitate deliberation by restricting the value of information that is necessary and sufficient for the completion of a particular task. As Nozick argues, the efficient completion of many common daily tasks requires the capacity to employ heuristics that are specifically adapted to representing the properties of phenomena which are invariably needed in a given environment.[441] In this respect, the heuristics employed in guiding moral deliberation are similarly attuned to evaluating available information. Therefore, even though moral principles are typically quite general, this apparent indeterminacy in their application need not be considered defective, as such principles are designed to capture information that is likely to be *invariably* salient to resolving moral problems in many different settings. Thus, a general proscription against lying is justified by its corrosive effects upon the trust which is essential to maintaining co-operation in any social setting.

Insofar as lying often misrepresents or omits information that would otherwise be salient for assessing moral consequences, then 'not being lied to' can be regarded as a fundamentally pragmatic value in any moral context.[442] The empirical fact that exceptions to the rule are nevertheless tolerated or even praiseworthy in certain circumstances, only shows the acknowledged salience of other environmental factors in mitigating or suspending that rule, especially when the disclosure of full and accurate information is likely to have clearly *immoral* consequences. Typically, when such consequences are likely to result in a foreseeable and irreparable injustice or severe harm, as would be the case for Jews living under the Nazi regime, then the moral salience is reversed so that lying to the Nazis is not only permissible but obligatory on the strongest grounds. Of course, even on the most trivial occasions, where insignificant information is nevertheless likely to cause offence or otherwise be given a disproportionate salience by others, then such considerations also function as pragmatic norms for restricting the general moral value afforded by truthful communication. But whether it is a case of lying to the Gestapo to save innocent lives, or lying to one's spouse about some minor imperfection, such exceptions are normatively justifiable by the same primary *social* value in upholding the ethical principle *against* lying generally.

Accordingly, as variations in the ethically salient aspects of social circumstances serve as *objective* grounds for applying different heuristics specifically adapted to recognise those ecological aspects, then the *evolutionary stability* of those heuristics is in fact a *necessary* condition for the continuing moral consciousness of agents. Without a common grasp of such principles, individuals would be deprived of the power to identify and respond to each other's interests and needs. As noted earlier, this is precisely where autistic individuals suffer substantial restrictions, as the development of successful prospective agency is dependent upon a certain proficiency in interpreting and predicting the behaviour of others in different environments. In this sense, learning to detect the salient contexts in which various moral principles are normatively applied is a process akin to language acquisition. By learning to identify the range of contexts in which a word obtains a specific reference, an individual is able to share information that is both necessary *and* sufficient to co-ordinate their actions. Nevertheless, in evolutionary terms a common language not only has the potential to reduce conflict and enhance mutual aid, but also facilitates strategic manipulation by deception.

In referring to cultural artifacts for their special importance in contributing to the agent's nonsubtractive and additive well-being, Gewirth fails to recognise that the substantive moral content afforded by such products cannot be sufficient to exclude their use for other amoral or immoral purposes. In fact, this insufficiency in fixing an *exclusively* moral value to any specific cultural practice or personal attribute is a defining feature of moral deliberation itself. After all, if the substantive moral value of any prospective action were at once sufficiently determinable by recourse to some infallible principle or procedure, there would no longer be any need for moral deliberation. Deliberation is required only because moral judgements are *not* sufficiently constrained by objective invariances or norms. Even acting in accordance with the most well-established moral principles still involves an act of deliberation insofar as the agent must first *rule out* the possible salience or priority of *alternative* principles in a given situation. As noted earlier, deliberation typically arises when a situation is experienced as *problematic*, in presenting some new aspect which may be more or less significant in its

relevance to some accepted principle or norm. Therefore, the intuitive and principled regard for human life need never be called into question, *unless* it comes into conflict with *other* principles or *other* human lives, as in the case of self-defence in war, abortion or other emerging issues in bioethics. In defending "the need for constant revision and expansion of moral knowledge", Dewey again stresses the significance of newly emerging facts, so that "At any moment conceptions which once seemed to belong exclusively to the biological or physical realm may assume some moral import."[443] For precisely this reason, no moral principle is beyond revision in the extent to which new factors may contribute to the range and depth of its of application to various personal and social issues. For example, while surgical techniques continue to improve the prospects for successfully separating conjoined twins (even those whose brains are partially fused), their parents must still confront the real possibility that one or both twins may not survive the operation. On the other hand, if the operation is not attempted, there may be an even greater probability that neither child will survive into adulthood. While extremely rare and distressing, such dilemmas confirm Dewey's contention that "A moral principle, then, is not a command to act or forbear acting in a given way: *it is a tool for analyzing a special situation*, the right or wrong being determined by the situation in its entirety, and not by the rule as such."[444]

The evolution of language is also analogous in this regard, as the pronunciation and semantic references contained in any person's working vocabulary must always permit some latitude in diverging from any fixed conventions of time and place. Insofar as the stability of such conventions is only the result of a selection process that may be more or less intelligently modified by the combined decisions of individual speakers and writers, then the meaning of a word can only ever be defined by its *use*. However, as the evolution of linguistic meaning cannot be a private affair, as it necessarily involves at least two parties with each one engaged in the dual roles of sender and receiver, no single individual can dictate any definitive word usage in any context.[445]

As suggested by the expression "to coin a phrase", the process by which a new semantic reference gains currency is determined by selection principles which have themselves been aptly phrased as "replicator dynamics".[446] By such principles, the adaptive fitness of the new entry will be determined by a number of factors, most notably those of expressive economy and mnemonic facility. Whether it be a minor change in the accented pronunciation of a rarely used word, or a distinct lexicon imported directly from another domain or language, the novelty's survival prospects will tend to reflect its utility within a framework of conventions in a particular social environment. This process need not involve any deliberative directionality on the part of agents, but may well proceed in a largely spontaneous sampling by minimal trial and error or sheer impulsive imitation. As a result of such scant attention, many changes in usage are analogous to genetic mutations, first emerging as a result of an error in copying information. For instance, in recent years among younger speakers the term 'issue' has developed a new meaning synonymous with 'problem', as expressed in the sentence "he has emotional issues". In this case, the rule denoting the *conventionally restricted context* in which 'issue' and 'problem' are rendered synonymous has probably not been adequately grasped, due to a deficiency in exposure to a wider range of conversational norms. Unless those norms are strictly adhered to in the population at large, the new usage may rapidly gain ascendancy in the emerging generation of

speakers, as a result of strong selection pressures inducing imitation and conformity at the expense of intelligent control.

This description of the mindless passivity of human behaviour in response to evolutionary pressures has ironically been popularised by certain proponents of Dawkins' 'selfish gene' theory, by which cultural products or 'memes', including language, are conceived as elaborate vehicles whose fitness is ultimately determined by their facility in replicating genetic information.[447] While some imitative cultural behaviour may well be explicable in such terms, particularly when it is received with relatively immediate biological benefits, the same behaviour will be subject to the more proximate selection pressures imposed by rewards and punishments of prevailing political and moral conventions, which work to *suppress* those self-seeking actions which have considerable costs for the community. While in agreement with Huxley on this antithetical distinction, even those like Dawkins and Williams who focus on genes as the ultimate unit of replication, whether via natural or cultural selection, also note the *positive* value in being *cognisant* of the inherited mechanisms which evolved to promote the survival of genetic information at the expense of ethical goals.[448] As Singer argues, even if the disposition toward mutual aid evolved primarily for its benefit to individuals or genes, that disposition can be intelligently cultivated and directed towards the consistent performance of genuinely ethical goals, thereby diminishing the value of selfish competition.[449]

In their criticisms, both Huxley and Gewirth fail to address the *evolutionary* value of the capacity for *critical reflection* as a countervailing process of organic *rational* selection directed by the behaviour of autonomous agents themselves. As a deliberative capacity that allows the organism to *learn* from mistakes, instead of thoughtlessly replicating them, it is essentially a heuristical method of information processing which is a precondition for cultural evolution. However, as the organic selection of cultural products reshapes the ecological affordances of nature, that process contributes to the course of natural selection itself. Therefore, the artificial selection of ethical principles cannot be causally independent of the cosmic or biological process, as human behaviour is itself an agent of ecological variation, which in turn, conditions the range and content of biological affordances favourable to the further development and maintenance of human agency. For this reason, cultural artifacts are themselves inherently biological in serving the physical and psychological well-being of the human agents who construct them. But as their well-being is also dependent upon the inherited skills and capacities which enable adaptation to evolving ecological conditions, the moral and political systems that agents construct must proceed from these biological constraints and affordances.

Learning to selectively adapt inherited heuristics and norms to a diversity of tasks and conditions would be extremely problematic if such phenotypic plasticity were reducible to genetic selection mechanisms. Ethical behaviour can only be learnt and applied at a level of consciousness that cannot be genetically assimilated or transmitted by specific algorithms which lack the flexibility afforded by general heuristics.[450] Implicit in the development of moral consciousness and action is the ability to recognise and respond to a *deficiency* in their operation, which can only be discovered by reflecting upon the ethical implications of following a particular norm or convention. Therefore, even though evolutionary stable strategies of normative behaviour can be selected without such a level of conscious reflection by individual agents, for that very reason the results of such a process would not

be produced by any mutual intentions to agree upon fixing their conditional preferences.[451] As Lewis suggests, preferences that arise by mutual imitation express an unconditional accommodation to the behaviour of others which does not qualify as a genuine convention where "…each wants to conform if the others do, and each wants the others to conform if he does.".[452] Thus, while the replication and exchange of information remains a necessary condition for the generation of normative rules which facilitate co-operation, the evolution and preservation of *prospective*, *purposive* agency is dependent upon each individual's capacity to exercise control in determining the grounds for following a specific rule or practice in a given context. Accordingly, the rational selection of ethical principles is premissed upon the *dynamic* power of deliberative agency in adapting those principles to respond to variations produced as a consequence of their prescribed application.

The phenomenon of 'political correctness' provides a particularly ironic illustration of the deliberative incoherence that can be generated, when organic selection dynamics are neglected in a misguided attempt to prescribe a fixed content to a range of possible representations in a given domain. While initially emerging as a well-intentioned policy typically aimed at rectifying some socially prejudiced imbalance, such as the under-representation of women in certain professions, the bias has all too often been 'corrected' by formally instituting countermeasures which restore equality only by being as *equally* unjust, ignorant or petty-minded as the original practice. Consider the evolution of the current academic convention where the use of the pronoun "she" has gradually supplanted the earlier practice where "he" was undoubtedly dominant in the literary tradition, perhaps simply reflecting the disproportionate representation of men employed in academia. For this very reason the choice of the male pronoun by any individual academic need not always reflect any personal or conspiratorial bias against the prospective employment of women, but may sometimes be intended as a *un*biased, factual account of the contemporary proportion of men encountered in that person's social environment. However, in the absence of this possible excuse it is undoubtedly true that "the exclusive use of 'he' to refer to both women and men is open to the charge of some implicit sexism;" as Sen observes. Given that representative fairness is the ethical rationale for abiding by the 'she' convention, then by that same logic its *conventional* use renders it *unethical*, as it fails to acknowledge the presence of every male academic, as Sen also recognises in remarking that "the exclusive use of 'she' can sound somewhat self-conscious and precious (and also open to a similar charge of discrimination from the opposite direction);".[453]

Given that the academic use of either pronoun in philosophy most often refers to no identifiable person, living or dead, but some fictional cast of characters in a possible world, then it is difficult to imagine how a non-existent *Jack* or *Jill*, *A* or *B* could feel aggrieved at their biased representation. The persistence of the 'she' convention, in spite of its incoherence, shows how easily education and critical reflection can be overwhelmed by the pressures of cultural conformity, leading to the fixation of a practice which effectively inhibits autonomous deliberation by default. In this example, the consequences are trivial and amusing, but if the same uncritical disposition is allowed to direct decision-making in other social areas, the power of personal agency can devolve into habitual *akrasia*, where responsibility for one's choices can be artificially selected by following the standard conventions associated with each specific activity. Therefore, contrary to Gewirth's claim, the

'problem of specificity' does not consist in any lack of determinate opportunities for evolving ethical thought and action, but in exactly the opposite threat where heuristics designed to guide reflection and action evolve into stereotypical responses which now come to function like *deterministic* algorithms. Under the weight of conventions which discourage reflection, moral prescriptions originally and rightly conceived as hypothetical imperatives, exercising freedom *in order to specify* the agent's considered choice of means and ends, are liable to appear as categorically specified by current, local practices and commodities.

A graphic example of the workings of organic selection in relation to moral agency can be witnessed in the following case. In affluent, urban areas parents with young children have rapidly cultivated a normative preference in selecting the 4-wheel drive as the family car. While the vehicle is the product of a technological culture, it has a primary biological function in affording protection from the elements, vastly expanding the prospects for discovering further habitable regions, natural resources and food sources, saving time and energy in transporting people and goods etc. Though purpose-built to survive rough encounters in the wild, the flexibility of its 'off-road' design allows for 'on-road' excursions into urban areas, testing its adaptive fitness for purposes afforded by those urban conditions. Here, through human agency the landscape has been pragmatically restructured and dynamically planned to accommodate the proliferation of means and ends which express and enhance that common agency. Thus, while the roads are sufficiently wide to accommodate larger vehicles like semi-trailers and 4WDs, they are not ideally suited to such conditions. As the road rules and traffic signals function as conventions to minimise collisions and casualties, they are implicitly directed by a moral concern for the prospective freedom and well-being of all road users, including pedestrians. Of course, in using the 4WD out of concern for their children's safety, the parents are undoubtedly motivated by the most primitive and praiseworthy familial concerns. But if the wider, long-term social consequences of that strategic behaviour are not given due consideration, then it is bound to proliferate, degenerating from a moral to a prudential mode of self-assertion, as more and more parents are forced to upgrade their vehicles to protect *their* children from this new menace. Meanwhile, those who cannot afford such vehicles are faced with the ever increasing prospect of being severely injured or killed from the impact of the ubiquitous 4WD bullbars which are, of course, originally designed to *protect* human lives. The myopic individualism fostered by such conditions is further compounded in this scenario, as the urgency in attending to the family's growing safety concerns prevents the parents from giving any serious thought to the long-term deterioration of the physical environment produced by the proliferation of these pollutant, gas-guzzling vehicles. Ironically, the very concern which precipitated the strategic choice of vehicles may thus become responsible for its subsequent neglect, as the parents' own children come to inherit an environment that is significantly *less* conducive to their well-being as adults.

As this example shows, it is not the selection mechanisms themselves which are unable to 'deliver the ethical goods', so to speak. When those mechanisms work to reinforce self-assertion at the expense of ethical constraints, they are simply reproducing a tendency in thought and action which must already be present to some degree in the population at large. If the child rearing instinct can lead some parents to knowingly discount the interests of others, then it is the lack of ethical content selected

by *their* thought processes which is the source of the problem identified by both Gewirth and Huxley. However, it is exactly the emotional strength and reliability of this instinct which also makes it indispensable for developing a similar care and attention to the needs of others. Responding to Huxley's pessimistic analysis, Dewey notes:[454]

> That self-assertion which we may call life is not only negatively, but positively a factor in the ethical process. What are courage, persistence, patience, enterprise, initiation, but forms of the self-assertion of those impulses which make up the life process? So much, I suppose, all would grant; but are temperance, chastity, benevolence, self-sacrifice itself, any less forms of self-assertion? Is not more, rather than less strength, involved in their exercise? Does the man who definitely and resolutely sets about obtaining some needed reform and with reference to that need sacrifices all the common comforts and luxuries of life, even for the time being social approval and reputation, fail in the exercise of self-assertion?

In this light, what defines a disposition as a virtue or a vice is not any intrinsic quality, but a reflection of the purposes for which the agent selects them as means. Thus, as Foot argues, courage may be equally required to murder someone for gain, as it is for achieving some moral end.[455] As it essentially consists in the capacity to suppress fearful instincts in the face of some perceived danger, it is a power whose assertion is vital to the completion of many tasks. But even traits like prudence or temperance can sometimes operate as habitual defects of character, the former issuing in "an over-anxious concern for safety and propriety", the latter associated with "timidity or with a grudging attitude to the acceptance of good things".[456] Like the freedom and well-being that Gewirth rightly defends as necessary and sufficient for *any* prospective purpose conceived by the agent, so too many character traits co-evolved for their vital contribution to the self-assertion expressed in any purposive action. But the power of self-assertion is also critically expressed in the *willpower* needed to exercise the *self-control* that allows autonomous agents to recognise and respond to the morally relevant aspects of a given situation.

Given these dynamic conditions, the PPA who reasons solely in accordance with Gewirth's generic and static conception of freedom and well-being is deprived of the capacity to strategically locate and control the fluctuating affordability of those goods for the completion of specific tasks in distinct environments. Gewirth's dialectical procedure and the logic of universalisation do indeed establish the categorical consistency that requires each PPA to exercise their freedom only insofar as it does not remove the conditions minimally required for the effective agency of other PPAs. However, the *internal* dialectical necessity that logically commits each individual PPA to accept the freedom and well-being claimed by all other PPAs is not, by itself, sufficient to generate any more than a universal right for all such agents to "control their behaviour by their unforced choices".[457] Unfortunately, freedom of choice in itself, even when supported by the material and psychological well-being needed to sustain its continual efficacy, cannot even begin to assess the moral content of any prospective purpose. The specific moral value of any purpose can only be determined from a perspective of agency which is informed by the ecological and dynamic constraints and affordances produced by the *external*

dialectical process of organic selection described above. Moreover, as full prospective agency can only be achieved and maintained by exercising intelligent and strategic control in adapting one's choices to those evolving conditions, it is this informed capacity for enhanced *autonomy* rather than Gewirth's generic freedom of choice which must be secured. Nevertheless, far from undermining the validity of Gewirth's rational criteria and procedure for establishing the fundamental rights of agency, these wider evolutionary modifications will hopefully succeed in fortifying many of his conclusions and recommendations.

---

[424] Thagard, op. cit., pp.72-73

[425] DePaul, op. cit., pp.28-30

[426] Thagard, op. cit., p.42

[427] Ibid., pp.76-77

[428] Stephen P. Stich, "A Pragmatic Account of Cognitive Evaluation," in *Naturalizing Epistemology*, ed. H. Kornblith (Cambridge, Mass ; London, MIT Press, 1994), pp.418-419

[429] W. V. O. Quine & J. S. Ullian *The Web of Belief* (New York, Random House, 1970), p.6

[430] W. V. O. Quine, "Natural Kinds," in *Naturalizing Epistemology*, p.66

[431] Alan Gewirth, "The Problem of Specificity," in *Biology and Philosophy*, 1 1986 p.301

[432] Thomas Huxley, "Evolution and Ethics," in T. H. Huxley and J. Huxley, *Evolution and Ethics, 1893-1943* (London, Pilot Press, 1947), p.51

[433] Ibid., pp.38-42

[434] Ibid., p.39

[435] Alan Gewirth, "How Ethical Is Evolutionary Ethics?" in *Evolutionary Ethics*, ed. M. H. Nitecki & D. V. Nitecki (Albany, SUNY Press, 1993), pp.245-246

[436] Ibid., p.253

[437] Reed, *Encountering the World*, p.149

[438] Ibid.

[439] Philippa Foot, in "Morality as a System of Hypothetical Imperatives," pp.160ff

[440] Radu J. Bogdan, *Interpreting Minds: The Evolution of a Practice* (Cambridge, Mass., MIT Press, 1997), p.239

[441] Robert Nozick, *Invariances: The Structure of the Objective World*. (Cambridge, Mass., Harvard University Press, 2001), pp.99-100

[442] Gewirth, "How Ethical Is Evolutionary Ethics?", p.252

[443] John Dewey, "Moral Judgment and Knowledge," in *The Essential Dewey*, vol. 2, p.339

[444] Ibid., p.338

[445] Cf. Wittgenstein's discussion in op. cit., § 41 and § 269 ff.

[446] Skyrms, op. cit., p.11 n.21

[447] See Dawkins, *The Selfish Gene*, pp.206-15; Susan Blackmore, *The Meme Machine* (Oxford, Oxford University Press, 1999).

[448] George C. Williams, "A Sociobiological Expansion of *Evolution and Ethics*," in Paradis, J., G. C. Williams, Et Al. (1989). *Evolution & Ethics: T. H. Huxley's Evolution and Ethics with New Essays on its Victorian and Sociobiological Context*. Princeton; Guildford, Princeton University Press, p.214

---

[449] Singer, *The Expanding Circle*, pp.168-169

[450] See Robert Nozick's distinction in *The Nature of Rationality* (Princeton, Princeton University Press, 1993), pp.166-167 & 173 and re ethical norms in *Invariances: The Structure of the Objective World* (Cambridge, Mass., Belknap Press of Harvard University Press, 2001), pp.243 & 276

[451] David Lewis, *Convention: A Philosophical Study* (Oxford, Basil Blackwell, 1986), p.119

[452] Ibid., p.120

[453] Amartya Sen, *Inequality Reexamined* (Cambridge, Mass., Harvard University Press, 1992), p.xiv

[454] John Dewey, "Evolution and Ethics," in *Evolutionary Ethics*, pp.101-102

[455] Foot, "Virtues and Vices," in *Virtues and Vices and Other Essays in Moral Philosophy*, pp.15-16

[456] Ibid., p.18

[457] Gewirth, *Reason and Morality*, p.59

CHAPTER EIGHT

DIMENSIONS OF AN ORGANIC SOCIAL CONTRACT:
THE MUTUAL SELECTION OF SOCIAL VALUES

(i)   Adaptive Moral Agency

The preceding discussion has concluded that the evolutionary conditions of human agency largely substantiate Gewirth's rational defence of freedom and well-being as goods that must be intrinsically[458] valued by all prospective, purposive agents. But the ecological and dynamic dimensions of those conditions have been shown to impose pragmatic constraints upon the degree to which such basic goods may be coherently evaluated as generically good for any particular purpose. Gewirth's definitive recognition of the prospective direction of rational agency does, at least, provide a formal basis for addressing the dynamic coherence of the PPA's evaluative system. Also, in acknowledging the degrees of necessity by which freedom and well-being vary in their contribution toward different purposes, Gewirth does indeed refine his generic conception of agency in its bearing upon specific purposes. However, as the latter concession thus weakens his claim to establish the PPA's generic rights to the *same* degree of freedom and well-being, irrespective of their purposes, Gewirth's Principle of Generic Consistency has limited force. Though it does establish a categorical right to a level of well-being minimally required to sustain the agent's prospective freedom of choice, the generality of its scope for any individual PPA's practical purposes is limited only insofar as the pursuit of those purposes does not thereby infringe upon the same right of other PPAs to effectively pursue their purposes.[459] While that singular dialectical restriction on the liberty of each agent's purposive actions

can certainly succeed in prohibiting a range of behaviours injurious to the minimal conditions of agency, it fails to provide for the evolutionary conditions of *moral* agency which requires a level of autonomy responsive to the organic dimensions of deliberation and action.

The level of autonomy sufficient for moral reflection and action accords with Haworth's definition of 'normal autonomy' described earlier. While minimal and normal autonomy both require a measure of deliberative competence as a condition for the development of self-control, the transition to normal autonomy is distinguished by the attainment of procedural independence in guiding one's substantive decisions.[460] Applied to the earlier analysis of behavioural conventions, minimal autonomy is evidenced by the agent's procedural dependence in uncritically following others' substantive judgements. By contrast, normal autonomy exhibits a capacity for independent reflection upon the substantive value of received ideas, norms or conventions. Therefore, even if the results of such autonomous deliberations should nevertheless lead an agent to fully endorse an entire set of traditional moral beliefs or practices, that commitment will have been guided by a rational assessment of its conceivable implications for that agent's prospective purposes. As long as a commitment does not also entail the foreseeable loss of that independent power of reflection in guiding each individual agent's progressive reassessment of their personal well-being, it may well serve the mutual interests and needs of many agents to bring their actions into conformity with certain established moral conventions. Thus, an individual's substantive dependence upon such conventions need not indicate a failure to independently assess their moral worth, as that person's commitment may, in fact, follow precisely from their own independent reflections. In this respect, Haworth suggests that even in the life of a cloistered nun, her substantive dependence upon a strict regimen of prescribed activities may be the result of the most serious and sustained personal reflection on the virtues of such a life.[461] Therefore, even though her daily life is governed by rules and practices which she has not personally devised, her autonomous evaluation of their substantive purpose within that religious life ensures that her dependence upon those principles is intentionally and rationally self-imposed. However, unless she is already naturally disposed to leading an ascetic life, her commitment to those daily tasks will only be regularly achieved through some measure of self-discipline. This is where self-assertion shows its vital function in enabling the agent's higher order considered judgments to exercise authority over first order impulses and received ideas.

Depending on the range of the agent's particular social interests and needs, the conventions observed in various domains will be more or less acceptable, by the degree to which they sustain the development and exercise of the level of autonomous self-reflection described above. As that self-control is a prerequisite for successfully engaging in the social interactions which facilitate the completion of common goals, all PPAs are already substantively dependent on the autonomous decisions of other PPAs in a given social sphere. By intentionally reproducing and adapting norms and conventions to accommodate variations in the range and content of available means and ends, the co-operative efforts and critical moral consciousness of such agents sustains the normal level of autonomy in a population. Therefore, in defining its evolutionary function, it is perhaps best described as *adaptive moral agency*. The value of any given heuristic, norm or convention can henceforth be broadly assessed by its prospects for advancing the adaptive moral agency of all PPAs.

In spite of the breadth of this positive characterisation of moral agency, it may still seem vulnerable to the organic selection pressures outlined earlier, whereby conventions often have the opposite effect in discouraging independent thought and action. Yet, it is only by learning to recognise and anticipate the environmental conditions which produce such pressures that adaptive agents can avoid unwillingly and unwittingly succumbing to their influence. For that very reason, the task of assiduously adapting local and domain-specific conventions to meet the universal conditions of moral agency is too onerous and important to be left to the voluntary efforts of individual agents. As illustrated in the 4WD scenario, in competitive conditions there is a pervasive tendency for wide moral reflection to yield to the pressures of prudential self-preservation, resulting in a collapse in the stability of conditions needed to sustain moral agency. In the previous chapter, it was concluded that the biological necessity of these cultural conditions warrants the PPA's right to claim them as vital for evolving the moral reflection and action which assures their social welfare. Implicit within the concept of a right is, of course, the recognition that individuals cannot personally ensure that they will not suffer some irredeemable loss or injury as a result of the behaviour of other agents. Whether through malice, negligence or wilful ignorance, in any society there is bound to be a proportion of rational agents with the power of normal autonomy, who nevertheless freely choose to exercise that power to repeatedly infringe upon the good will of moral agents.

(ii)   Equality, Rationality and Reasonability

As described earlier, it is this perennial problem of the amoralist or 'free rider' that most likely led to the evolution of the social contract, as a political device for generating a public appreciation of the common good that must be protected by maintaining stable relations of mutual trust and dependence. It is in response to this relative instability that the political sense of justice emerges. As Rawls notes: "To insure stability men must have a sense of justice or a concern for those who would be disadvantaged by their defection, preferably both. When these sentiments are sufficiently strong to overrule the temptations to violate the rules, just schemes are stable".[462] In this connection, Rawls also acknowledges Trivers' account of the evolution of reciprocal altruism, in showing "how the system of the moral feelings might evolve as inclinations supporting the natural duties and as stabilizing mechanisms for just schemes".[463] But even with such schemes in place, Rawls agrees that individuals need the protection of government to effectively ensure their enforcement "even when everyone is moved by the same sense of justice".[464] Accordingly, in Rawls terminology, a "well-ordered" society, conceived as a "fair system of co-operation" is sustained by a public conception of justice, which works to ensure that the terms of such co-operation are not only mutually advantageous, but *equally* so for all citizens.[465] The notion of equality is thus central to the basic structure of a fair system of reciprocity, in preventing the evolution of asymmetrical relations of mutual dependence, whereby the effective agency of one or more parties is enhanced at the cost of a corresponding diminution of that power incurred by other parties. Though Rawls willingly concedes that this political conception of "justice as fairness" also incorporates a fundamentally *moral* conception inherited in a democratic culture, he argues that it is nevertheless intuitively acceptable as a firm basis from which reflective

equilibrium may proceed in representing citizens from that culture.[466] The critical weakness of this overly conservative approach to the use of reflective equilibrium has already been exposed as grossly irresponsible from both a psychological and philosophical perspective, at least insofar as these disciplines aim at rectifying inherited biases and errors of judgement. Furthermore, by restricting his conception of justice to the equality of reciprocal relations in democratic societies, Rawls only devalues its universal basis in the evolution of all surviving moral systems. Although in the *Law of Peoples*[467] Rawls does attempt to extend this universal notion of fairness to defend a global account of justice, his constructivist approach to justifying its foundational status, both politically and morally, would be on much firmer ground by proceeding from an evolutionary interpretation of its cultural function.

In citing the contractualist merits of Trivers' evolutionary account of reciprocal altruism noted above, Rawls was only concerned with "simply checking whether the conception already adopted is a feasible one and not so unstable that some other choice might be better".[468] As Trivers' hypothesis had only just been published in the same year as this reference in Rawls' *A Theory of Justice*, its brief and limited consideration is certainly judicious in that context. However, like Darwin's theory itself, Trivers' hypothesis has since been firmly validated, so that it can serve as an empirically plausible basis for modelling an evolutionarily stable social contract that is compatible with all human cultures. In contrast to the fragility of Rawls' 'intuitive' defence of its acceptability in one cultural tradition, an empirical account of the universal acceptability of justice as fairness is fully consistent with the principles of explanatory coherence specified earlier, where reliable evidence has a special priority in affirming the feasibility of a hypothetical belief or policy. Of course, the historical stability of reciprocal altruism does not, in itself, suffice to warrant its unaltered normative acceptance in governing social relations, as there may in future be some evolving prospects for significant improvements in its functional coherence. However, it is especially noteworthy for highlighting the inherently *epistemic* function of the concept of equality in maintaining the stability of reciprocal relations among free agents. Given its role in enabling individual agents to measure and recall disparities in the mutual benefits and contributions realized by their interactions, equality in social exchange remains first and foremost a principle of pragmatic quantification that is essential for the detection of free riders in their various incarnations. This epistemic rationality implicit in the political and moral value of equality has already been noted in defending the principle of universalisability against the objection that it is premised upon an inherently moral conception of equality, in judging the significance of relevant similarities among prescriptions. Once again, an evolutionary analysis of the concept shows its evaluative function to be universally pragmatic in enabling agents to avoid suffering undue losses in the goods needed to sustain their prospective agency. As such, the epistemic function of equality in free social transactions remains transparently and unequivocally basic in the construction of any viable moral and political system. Accordingly, both these *evolved* pragmatic and epistemic *constraints* inherent in reciprocal altruism and its robust stability in *continually sustaining* the mutual autonomy and well-being of all prospective, purposive agents restricts the sustainability of any alternative system which does not support these evolving agency needs to the same degree.

This instrumental, quantitative value of equality in determining and monitoring the fair terms of co-operation does not, of course, exhaust or reduce the significance of the moral conception of

equality in its application to the powers of normal agency. In democratic cultures, equality not only represents an external measure of reciprocal relations among agents, but also serves to denote a conception of the irreducible individuality of the agent *per se*. As Dewey observes, this moral understanding of the uniqueness of the individual is at odds with the strictly quantitative conception, whether expressed in the hierarchies of "rigid feudalism", or in the "atomistic individualism" of traditional democratic culture. In these moral terms, "…equality does not mean mathematical equivalence. It means rather the inapplicability of considerations of greater and less, superior and inferior".[469] However, as Rawls later came to realise, the intuitive appeal of democratically informed "comprehensive doctrines" of moral personality, such as those elaborated by Kant and Mill, cannot be allowed to dictate the terms of reciprocity acceptable to citizens concerned with determining the grounds of justice for a society informed by a "plurality of reasonable doctrines".[470] Instead, Rawls' condition of reasonability is designed to ensure that all citizens may accept principles of *political* justice which represent them "as free and equal, and not as dominated or manipulated, or under the pressure of an inferior political or social position".[471]

Rawls' notion of reasonability is thus implicitly connected to his description of the essential moral powers of agents i.e. a sense of justice and a conception of the good.[472] As such, a shared disposition for moral motivation is already assumed to function with equal potential in the psychology of citizens in a well-ordered society. As long as that potential can be developed to approximate a common threshold of proficiency, agents can be considered as equal in their possession of those powers. Therefore, even though advantages in natural endowments or social conditions will enable some agents to be considerably more proficient or conscientious in the use of these powers, Rawls rightly notes that such contingencies have no bearing upon the *common* goal of determining principles of justice as fairness which can be reasonably accepted by any agent in sufficient possession of such powers.[473] Reasonability thereby sets a normative standard of mutual understanding and responsibility which need not be derived from, or committed to any substantive moral doctrine, as it is only concerned with the political task of establishing the fair terms of social co-operation among citizens espousing diverse moral conceptions or commitments. Therefore, while the principle of fairness suffices to ensure that the agreed political conception of justice must also be a moral conception, there need not be any question of circularity here. Insofar as *any* moral perspective, by definition, must take sufficient account of the interests of others, then the notion of reasonability simply sets that standard of deliberation so that, for instance, as Scanlon argues, it is "reasonable to object to principles that favor others arbitrarily".[474] The principle of impartiality explicitly serves to *invalidate* any attempt to derive a political conception of justice from a substantive moral conception representing the interests of some members more than others. As such, the impartiality assured under a social contract effectively *prevents* the potential circularity of permitting an existing moral system any political self-justification. It is in this sense that Rawls defines the political conception of justice as a "freestanding" view, given its conceptual independence from any particular moral standpoint, while still remaining an essential component of any reasonable comprehensive doctrine.[475]

Reasonable agents are thus distinguished from those that are merely rational, by their willingness to both seek and abide by principles of justice that are mutually acceptable, or rather "could

not be reasonably rejected", as Scanlon suggests.[476] From the perspective of agents concerned to preserve their prospective autonomy and well-being, Scanlon's formulation has substantial merit in promoting a degree of mutual respect and worthy consideration of proposed principles, which are bound to include impartiality from the outset. By contrast, Rawls' attempt to artificially induce impartiality by imagining agents situated in an 'original position' behind a 'veil of ignorance' only prevents them from developing the sense of mutual trust and understanding which is indispensable for evolving moral agency in actual social environments.

Both Rawls and Scanlon explicitly appeal to the notion of reasonability for its dispositional value in motivating agents to adopt principles of fairness which would not be acceptable on purely rational grounds. To his credit, Rawls does not assume that rational agents are bound to be constantly or exclusively self-interested in their deliberations and in fact recognises agents' personal interests as a vital factor in motivating their desire for protection under a social contract. Indeed, in the 'original position' imaginary parties representing the interests of a specific group of citizens are fully expected to exercise their rational autonomy to best secure those interests.[477] In Rawls' scheme, both rationality and reasonability guide deliberations in distinct association with the agent's two moral powers: reasonability from a sense of justice, rationality on behalf of a conception of the good.[478] By contrast, Scanlon eschews the common view of rationality understood as "what conduces to the fulfillment of the agent's aims", regarding it as a misconceived account of the agent's deliberative task, which is essentially concerned with "how best to bring about the desired end of agreement on principles".[479] As his contractualism also eschews the artifices that Rawls employs to avoid any deliberative conflict between the agent's rational and reasonable nature, Scanlon resolves this potential conflict by excluding the possibility that rational choice, as ordinarily conceived, can motivate agents to agree on principles for deciding *moral* questions. Scanlon's dismissal of the role of rational choice in furnishing the grounds for contractual agreement is essentially based on the suspicion that "fulfillment of the agent's aims" can often only be realized at the expense of other agents' aims.[480]

While Scanlon's account of reasonability closely accords with the evolved norms of social and moral reasoning already discussed, that account need not be in conflict with the rational standards invoked by Gewirth and Hare. Though discounting Hare's test of universal acceptability and any other attempts to derive moral principles from logical criteria,[481] Scanlon argues that principles arbitrarily favouring some persons over others are to be rejected on grounds remarkably similar to those cited by Hare e.g. the irrelevance of proper names in conferring moral distinctions.[482] Similarly, in rejecting principles which would exempt particular individuals from the burdens of providing assistance to others, Scanlon notes that "…the generic reasons arising from the burdens that these principles involve for agents in general are not sufficient ground for rejecting a general requirement to aid (or not injure), given the reasons that others have for wanting this protection, and there is nothing special about my case".[483] However, this argument implies that the reasonability of such principles rests upon some *prior* justification for accepting the legitimacy of other agents' reasons for seeking such protection. In fact, it can only be reasonable *because* agents have purposes which they want to fulfill. As Gewirth argues from the criterion of relevant similarity, "Inconsistency is incurred if one accepts for oneself a rule of action one rejects for other persons".[484] As the relevant similarity here is simply that of being a

prospective, purposive agent, then all such agents have a necessary and sufficient reason to respect each other's need for protection, by the same *rational* principle of logical consistency. Thus, while reasonability is indeed required for each agent's rational autonomy to be effectively realized, such reasonability is itself derived from principles of rational consistency. After remarking that "It may not be possible to prove that the reasonable cannot be derived from the rational", Rawls claims that those attempts which may appear to succeed, do so only by introducing the notion of reasonability at some point. Rawls cites Gauthier's *Morals by Agreement* as the most notable "serious" attempt which nevertheless fails on precisely these grounds.[485] Gewirth praises Gauthier's theory for its realistic portrayal of the initial inequalities distributed among agents. But he also finds it implausible that the rational self-interest of Gauthier's agents could possibly lead them to accept binding moral constraints against their own interests.[486]

Before considering the evolutionary prospects of Gauthier's rational choice model, it is important to distinguish between different conceptions of mutuality and reciprocity among agents. Prior to determining the social conditions sufficient for the development of adaptive moral agency, it is essential to exclude any model of the environment which misrepresents the evolved capacities of agents and the inherited structure of social relations. On the basis of the preceding empirical investigations, it was found that the social contract governing all surviving stable societies is manifested in the reciprocity afforded by prevailing institutions and conventions. However, in spite of their evolutionary stability, the regulative structure of the contract still permits substantial inequalities in the normative terms of reciprocity, to the extent that even the most hierarchical societies may remain sufficiently 'well-ordered' while tolerating the most extreme disparities in the contributions exacted from various social groups. Therefore, even if reasonability sets conditions justifying a political conception of society as a 'fair system of co-operation', it is first vital to determine the extent to which reciprocity may serve to promote or inhibit the widespread development of adaptive moral agency, without begging the question as to what constitutes the fair terms of co-operation for such adaptive agents.

As already noted,[487] each adaptive moral agent (AMA) is logically committed to regarding every other AMA as free and equal in terms of their common potential for approaching a dispositional capacity for moral reflection sufficient to protect their own adaptive agency. However, the force of this logic is only binding upon agents who have already learnt to appreciate the necessary conditions of their agency in these evolving dimensions. Accordingly, it may seem that a right to the goods required for adaptive agency could readily be denied by PPAs who are able to successfully plan and pursue their purposes in an incremental, *ad hoc* fashion, without reflecting on the ecological or dynamic conditions which allow those purposes to find fulfillment. PPAs who are neither amoralists nor egoists may regulate their lives in accordance with precepts which deny the full potential of their own agency, or may be content to live relatively unobtrusive, solitary lives, so that they need never have occasion to consider their level of prospective agency as threatened or otherwise deficient in fulfilling their limited goals. Although agents such as Zeke and the obsessive piano player were shown to be genuinely incapacitated by denying themselves the goods required to develop an adaptive awareness and enhancement of their autonomy and well-being, their right to a level of goods sufficient for that minimal conception of their own agency, at least logically commits them to respecting all other agents'

equal rights to the same conditions needed for their minimal autonomy and well-being. Such myopic, maladaptive agents may even be more thoroughly in possession of Rawls' two moral powers than adaptive agents themselves, given that the latter have developed habits of critical reflection and cautious planning which are resistant to the ready assimilation of inherited beliefs and practices.

Members of an Amish community, for example, conduct their lives in accordance with a 'comprehensive moral doctrine' which discourages individual ambition and imagination, insofar as these tendencies conflict with their traditional ideals of asceticism and communal self-sufficiency. Therefore, individuals brought up in such a community would scarcely have cause to place a high priority on the scope of their personal autonomy, as it is largely consumed by efforts to preserve the community's autonomy. Though their communities are often partly sustained by some limited commerce with the wider society, the Amish have no good reason to engage in any long-term contractual reciprocation with their liberal neighbours, particularly where such contact is likely to expose them to beliefs and practices which are abhorrent to their own conception of the good. As DePaul argues, in order to make a rational choice among alternative moral conceptions, an Amish person should not have to experiment with them personally, especially when such behaviour can be suspected to "corrupt his judgement".[488] In virtue of this prudential self-concern, the Amish person also demonstrates his *reasonable rejection* of any imperative to adapt his behaviour to a cultural environment which threatens his moral integrity. Yet, in doing so, he ironically reveals his own dependence upon the *adaptive* power of moral agency, as shown by his recognition of the need to avoid experiences which can be expected to undermine his capacity for moral judgement. In effect, his reflections are, at least, partly guided by the same strategic principles of dynamic rationality which were cited in the paradigm case of Ulysses and the Sirens. In this case, the Amish person anticipates the possible attractions in a foreign environment that might tempt him to lose the moral powers which have thus far reliably served to sustain his well-being with the mutual support of his fellow community members. Of course, the reasonability of his assessment depends upon the reliability and extent of his knowledge regarding the potential affordances of that new environment, as well as its hazards. Having shown himself to be willing and able to carefully consider the adaptability of his prospective moral agency in a wider environment, his decision to remain in a stable social environment which he expects to meet his foreseeable needs is therefore an adaptive social choice.

The above example supports Rawls' claim that both rational self-interest and reasonability are sufficient to provide agents with the motivational capacity to agree upon the fair terms of co-operation, regardless of the distinct communities to which they belong. Furthermore, in highlighting the ecological and dynamic dimensions of the adaptive agent's prudential *and* moral reasoning, the agent's self-interest and moral interest are seen to coincide. In this case, in his capacity as a PPA, the agent's limited prospective purposes as a member of his Amish community could only warrant a right to protection of a level of autonomy and well-being sufficient to fulfill those purposes. As Gewirth has shown, on the grounds of logical consistency alone that same right would then necessarily apply to all PPAs, so that it would have to be unconditionally accepted by any rational agent. However, in recognising that his prospective agency is neither entirely exhausted, nor exclusively bounded by his community's resources or interests, the Amish person is forced to concede that he cannot reasonably

reject a claim to the extensive rights and institutions required for the development and practice of adaptive agency in the wider world. If he were to follow standard rational choice principles, where preferences are selected by their maximum potential to benefit the agent's short-term self-interested goals, he would not only fail to appreciate the benefits afforded to adaptive agents, but would also discount his dependence upon the adaptive capacities of his liberal neighbours. From such a myopic perspective, the most rational choice would be to reject principles that would tax his own community to support a statewide level of autonomy and well-being from which he could not expect to profit, given his much more limited purposes within the Amish community.

The distinctive social perspectives from which individual agents view their relations with others raises fundamental questions about the practical scope of their allegiance to a universal conception of a 'well-ordered' society as a fair system of co-operation. Various social commitments beginning with close loyalties to family, friends and colleagues, allow significant variation in the manner in which they are internally governed by different reciprocal norms. Such norms are bound to induce "strains of commitment" in their potential conflict with a political conception of reciprocity between agents regarded as free and equal citizens.[489] Rawls endorses Gibbard's view that this political conception "…lies between the idea of impartiality, which is altruistic (being moved by the general good), and the idea of mutual advantage understood as everyone being advantaged with respect to each person's present or expected future situation as things are".[490] While such a conception is remarkably consistent with the normative structure of reciprocal altruism, the evolved constraints of moral reasoning must also reflect the intrinsic flexibility of the epistemic concepts from which those norms evolve. As conditions for the evolution of human agency, norms of co-operation between parents and children can scarcely be conducted on the same free and equal terms expressed in Rawls' political conception of reciprocity. Even within the family unit, given the continual growth and variation in the degree of dependency between different members, the normative conduct of relations is necessarily dynamic and highly particularised in progressively adapting to individual needs. Accordingly, instead of prescribing fixed terms and conditions in its practice, reciprocity will tend to *fluctuate* between the governing ideas of impartiality and mutual advantage, as principles of equality and fairness are gradually and selectively learnt and adapted to different social contexts and purposes. Thus, insofar as the family as a social unit functions to advance the well-being of each of its members, the developing needs and capacities of particular individuals often demands the exercise of *partiality* in meeting their specific needs. However, from the perspective of the parents, their duty to provide for their children's welfare need not require equal reciprocation even when their children are in a position to provide it.

By contrast, once children have attained a certain level of autonomy, it is the *mutual* status of both parties as free and equal *agents* that protects the rights of each, rather than any reciprocal obligations. As Gewirth notes, this marks a crucial distinction in the rational basis for determining rights and obligations. While both mutuality and reciprocity describe "dynamic symmetrical relations" among agents, "In reciprocity, if A does good for B, then B does good for A. In the mutuality of human rights, A has rights against B and B has rights against A".[491] As Gewirth has shown, every PPA's prudential right to preserve their continuing agency follows directly from their constitutive dependence upon the goods necessary for purposive action. As such, the rational basis for that right is not

contingent upon the person having first undertaken or avoided some prior action which benefits one or other persons, who must now restore equality by reciprocating in some prescribed manner.[492] Nevertheless, the mutual necessity of the goods minimally required for any PPA's purposes suffices to establish equal rights to the fulfillment of purposes only insofar as those actions do not deprive other PPAs of that same level of basic goods. Therefore, notwithstanding the above distinctions, the mutuality in the basic needs of all PPAs can itself only be efficiently maintained if individuals regularly observe the same normative constraints in their interactions. In this sense, the idea of reciprocity is still operative in each agent's recognition that their respect for others' needs is returned by the same measure.

All PPAs share this mutual need for a minimal level of autonomy and well-being quite simply in order to retain their functional capacity for purposive action. Logical consistency alone requires each PPA to recognise every other PPA's similar need for such goods, thereby generating both negative and positive rights and obligations to safeguard each other's purposive capacities. However, in seeking to have those rights publicly recognised, endorsed and protected on the basis of these prudential concerns, such agents cannot reasonably reject moral principles which would recognise and develop each other's mutual potential for adaptive agency. Regardless of the distinct communities to which they belong, no prospective, purposive agent can truthfully claim their livelihood to be independent of the ideas and efforts of preceding and current generations of adaptive agents. In many respects, members of the strictest Amish community owe the stability of their prevailing mode of life to the ascetic dispositions inherited from their ancestors. In finding the resources to establish productive settlements on an alien continent, those early settlers demonstrated their capacity to strengthen their moral powers to a level capable of withstanding the constant tribulations of the new environment.[493] But the quality of life in any contemporary Amish community is undeniably dependent upon the good will and non-interference of those in neighbouring communities, while their rights to property and religious freedom are ensured by state and federal legislation produced by the democratic process.

Once again, principles of rational coherence can serve to test the legitimacy of substantive moral claims. In this case, the Amish person's rejection of liberal democratic principles is not simply unreasonable, but is deductively incoherent and based on false premises. Not only is the purported independence of his community demonstrably false, it in fact survives as a direct result of the robust liberal dispositions and practices of his neighbours and the democratic political institutions they uphold. As such, he cannot coherently maintain that his moral autonomy is threatened by a liberal, democratic conception of the good, when it is that very conception which continually supports and protects the political and moral autonomy of his community, while also ensuring his personal autonomy is protected from threats within the Amish community itself. In this respect, as the values and practices of that particular cult evolved from its roots in the Protestant culture of the Reformation, its adherents are also indebted to the spirit of those Enlightenment thinkers who challenged the legitimacy of clerical authority in determining questions of moral doctrine and conscience. Unlike most other Christian cults whose beliefs cohere with these Protestant foundations in valuing "…individual liberation from sin rather than submission to the corporate community of believers",[494] the strict conformism in Amish culture is inconsistent with the liberty of conscience which originally motivated

their formation and migration to a land *free* from religious authoritarianism. As Hardin notes, the Amish sect did not emerge from any uniform religious culture, but from a *variety* of beliefs among the original Mennonites. It was this plurality of perspectives and the recognition of the need for a radical revision of accepted beliefs which furnished the grounds for Jacob Amman's decision to establish a separate community.[495] Thus, wide reflective equilibrium and critical reflection from diverse perspectives prove to be indispensable conditions in the formation of even the strictest community, as it is dissatisfaction with existing regimes which precipitates the search for improvements. As such, it is fundamentally incoherent for members of such communities to deny the principles of individual self-expression, criticism and imagination from which their own moral autonomy arose. Indeed, this incoherence and intolerance of criticism and belief revision is threatening the survival of Amish communities today, as testified by the frequency of defections.[496]

From the extent to which any person's autonomy and well-being is historically and presently indebted to the adaptive agency of others, it logically follows that their prospective status is similarly indebted to future generations of adaptive agents. As Loewy succinctly points out "…we who exist today owe an obligation to the future because others whose future we were have in the past nurtured and sustained us".[497] As Hardin remarks, from the perspective of those cultures like the Inuit, whose traditional mode of subsistence is threatened by modern innovations, it is perfectly understandable that they should wish to preserve those traditions. However, like the Amish, scrupulous adherence to those cultural demands severely reduces the future prospects for their children, leaving them to suffer the burdens of surviving in a stagnant culture, rather than allowing them to receive a modern education which could only enhance their prospects.[498]

In spite of the dynamic causal responsibilities connecting present and future generations, the question of precisely how far such future concerns should extend can be expected to reveal substantial disagreement. But this *prima facie* disagreement on the extent and force of social obligations is one of the essential issues which the concept of an organic social contract is designed to address. Therefore, this question can be progressively considered and revised in accordance with ecological and dynamic principles of distributive justice, as they are rationally selected and applied using the same coherence criteria employed in the preceding analysis. From the universality of those standard principles in the evolution of human cognition, there can be no community of rational agents who can fail to be bound by those standards. Hence, there can be no justification for drawing artificial communitarian boundaries between agents who are all *equally* constrained by the same evolved norms and heuristics of practical reason and conditions of agency. For the same reason, a political conception of justice can neither be wholly derived from, nor restricted to agreement among moral agents inhabiting well-ordered societies which are abstractly depicted as *closed* "only because it enables us to focus on certain main questions free from distracting details".[499] Not surprisingly, among those distracting details are certain issues which create difficulties for Rawls' restricted account of the scope of rational agency and criteria for moral evaluation.

(iii)   The Reasonable Extension of Concern

Contractual restrictions in setting standards for qualifying as a recipient of moral concern are often cited as constituting serious if not fatal objections to the acceptability of most versions of contractarianism. These criticisms are based upon the assumption that those moral principles or norms which would be determined by hypothetical agreement among rational agents must therefore have limited force in creating obligations which are binding only between such agents themselves. However, in spite of the self-imposed tendency to construct models restricting the boundaries of agents' concerns to their own mutual interests, there are no *a priori* theoretical strictures which necessarily exclude the possibility of extending those boundaries. As noted above, parents' concern for their children's welfare is not normatively contingent upon the expectation of receiving some reciprocal benefit, as if the child's development were valued primarily as an investment toward the *parents'* future welfare or social status.[500] Therefore, even though contractualist models are logically bound to consider children as recipients of moral concern, given their normal potential to become active participants in reciprocating that concern, their potential moral agency need not function as the sole criterion for determining their status as worthy of consideration.

As Gewirth notes, the minimal level of rationality required for the exercise of prospective, purposive agency can be defined by the following practical functions: "the ability to control one's behaviour by one's unforced choice, to have knowledge of relevant circumstances, and to reflect on one's purposes".[501] As there are degrees to which these abilities may be approximated, some children may display cognitive deficits which permanently impair the development of their rational agency, in one or more of these three generic manifestations.[502] Yet, once such a child has been born into their care, few parents informed of such a prognosis would seriously consider abandoning their offspring, even before they have had occasion to be moved by a strong emotional attachment. Although certain cultures and ostensibly utopian theories[503] have recommended or condoned policies of abandoning or even killing disabled children and adults alike, whatever reasons are advanced to defend such practices invariably ignore the interests of those who are their victims. In the past, Darwin's theory itself has often either been grossly misinterpreted or spuriously represented as authorising social policies designed to promote the 'survival of the fittest',[504] thus providing a pseudo-scientific rationalisation for devaluing the existence of those who cannot survive without substantial assistance.

Even within the restrictions imposed by the most profoundly disabling mental or physical impairment, a person may still possess some partial or episodic capacity to reflect upon the circumstances relevant to their limited choices. This level of autonomy is comparable to that exhibited by other sentient creatures whose reflective consciousness has not evolved to a level that would allow the intentional formation of higher order beliefs and desires which could successfully prevail over instinctive or learned responses. But a deficiency in such cognitive capacities does not diminish the animal's capacity to experience pleasure or pain. On the contrary, as species equipped with well-developed nervous systems are all the more sensitive on that account, those who lack the strategic intelligence or physical skills to reduce their exposure to painful experiences are bound to suffer to a degree which no rational human being could regard as reasonable.

Many philosophers have long recognised the vital role of sympathy, or what is now defined as empathy, in providing an emotive disposition for moral action, as distinct from rational principles or

procedures which are held to be either incapable or insufficient to sustain such motivation. While Hume is often correctly cited as most representative of this tradition, his account of the process by which a person comes to sympathise with another's character or situation cannot avoid appealing to the person's *cognitive* grasp of the similarities which give *rational* relevance to the object of such sympathy. Hume observes that: "When any affection is infus'd by sympathy, it is at first known only by its effects, and by those external signs in the countenance and conversation which convey an idea of it….Accordingly, we find, that where, beside the general resemblance of our natures, there is any peculiar similarity in our manners, or character, or country, or language, it facilitates the sympathy."[505] Thus, in converting ideas into impressions "sympathy is exactly correspondent to the operations of our understanding;".[506] Hume also notes the workings of sympathy within various social species, whereby a range of emotions such as "fear, anger, courage" and even "grief" serve to communicate "passions".[507] Although anger, courage and grief imply some degree of reflective awareness which may only be partially attributable to such species, they at least describe the rudimentary emotions which are indeed relevantly similar to the expression of these human emotions. More importantly, as Darwin recognised, these emotions evolved and retain their essential epistemic function in signalling information.[508] Whether issuing defensive threats to competitors, or alerting conspecifics to various ecological hazards or resources, such sympathetic emotions thereby facilitate interactions for humans and other animals alike, both within and between species.

From this common heritage of sympathy, even without the same capacity for reflection, Hume notes that: "A dog naturally loves a man above his own species, and very commonly meets with a return of affection", so that "…by benefits or injuries we produce their love or hatred; and that by feeding and cherishing any animal, we quickly acquire his affections; as by beating and abusing him we never fail to draw on us his enmity and ill-will."[509] By comparison, Kant argues that a loyal dog who has grown too old to be of service to his master should not be shot, simply on the grounds that such cruelty may dispose the man to treat his fellow humans in like manner.[510] Bernstein cites Kant's narrow conception of contractualism as typifying the standard view, whereby other sentient beings attain moral considerability only insofar as they benefit the moral interests of humans.[511] Hume's reference to the dog's *return* of affection or 'ill-will' shows more regard for the extent to which reciprocity is not only possible but actually responsible for inducing characteristic behaviours in both the dog and his master. It is of no small importance to be reminded of the historical fact that the dog, as a species, is descended from wolves whose domestication by humans induced the selection of phenotypic traits especially adapted to co-operation and co-habitation with our species.[512] In addition to the more proximate survival advantages afforded to our ancestors by the dog's superior sense of smell, hunting and guarding skills etc. , the mutuality of this co-evolutionary process is manifested in a feed-back effect, whereby the dog's characteristic loyalty and protection promotes a correspondingly kind and appreciative disposition in its human companions.

The evolutionary dynamics in the relations between humans and other species are typically and extensively symbiotic, so that the undesirable side-effects of mutual dependence can also function as an irritant provoking remedial action which may well be advantageous to the evolving autonomy of man and beast alike. Thus, while 'man's best friend' may provide faithful service and watchful

protection, the dog introduces new parasitical pests such as fleas into the human environment. Kant himself shows an appreciation of such literal irritants, in remarking that:[513]

> The vermin that torment men in their clothes, their hair, or their beds, may be, according to a wise appointment of nature, a motive to cleanliness which is in itself an important means for the preservation of health. Or again the mosquitoes and other stinging insects that make the wilderness of America so oppressive to the savages, may be so many goads to activity for these primitive men, [inducing them] to drain the marshes and bring light into the forests which intercept every breath of air, and in this way, as well as by cultivating the soil, to make their habitations more healthy.

This latter observation is also particularly instructive in the context of the earlier discussion of America's pioneering religious communities, whose enduring subsistence can be attributed to a similarly adaptive use of available resources. Once again, by examining the operations of human agency in the dynamic and ecological context in which it evolves, it becomes clear that our own prospective agency is largely dependent upon the contributions of other animals. Furthermore, Kant is quite right in arguing that cruelty to animals encourages a similar disposition towards fellow humans. However, if the moral status of other animals is determined solely by their contribution to the moral development of humans, then their contractual enfranchisement is still only precariously maintained, as it is contingent upon their continuing serviceability for human interests. On this instrumental basis, a concern for protecting the well-being of various species could in fact be reasonably rejected if, for example, such protection would prevent their possible use as food for humans on occasions when other sources may be depleted.

The sympathetic understanding which engenders a propensity for mutual aid in the relations between humans and various domesticated species may well be sufficient to ensure the acceptability of long-term measures for safeguarding the welfare of those species. But this *conditional* justification for their contractual moral standing does not give due consideration to the important distinction between mutuality and reciprocity raised by Gewirth's argument for the sufficiency of agency, where it is the *inherent* mutual capacities shared by prospective, purposive agents which logically commits them to extend the same right of protection to all those minimally endowed with the same potential capacities. As such, it is not simply the agent's capacity to reciprocate which warrants moral consideration. Rather, it is the relevant similarities in the evolved capacities shared by all autonomous agents which provide the grounds for rational coherence in evaluating the moral status of different species. Therefore, while "the ability to control one's behaviour by one's unforced choice, to have knowledge of relevant circumstances, and to reflect on one's purposes" are conditions for approximating rational agency in humans, there is sufficient evidence to show that some other species also possess this abilities in "rudimentary forms", as Gewirth duly concedes.[514]

As argued in Chapter One, there is now increasing observational and experimental evidence to show that autonomous control, inductive reasoning and self-awareness are powers which have evolved by degrees in various species. This evidence alone is sufficient to refute Gauthier's somewhat *ad hoc*

attempt to exclude all non-human animals from moral consideration, on the assumption that the capacity for "semantic representation" is confined to human agents. Once again, the dog comes in for a beating in Gauthier's remarking that: "You can, as your dog on the whole cannot, represent a state of affairs to yourself, and consider in particular whether or not it is the case, and whether or not you would want it to be the case."[515] Though lacking the linguistic apparatus which facilitates the scope for semantic representation, a number of intelligent breeds are undeniably capable of both communicating their desires and reporting information to their human partners, which implies a level of adaptive cognition in learning to discriminate between various sources of ecological information in terms of their desirability for a specific task (e.g. hunting & sniffer dogs). As noted in Reed's summation of Gibson's research into perceptual cognition, ecological information is normally detected and processed *directly*, so that internal representations are superfluous in accounting for the practical intelligence of species lacking the capacity for semantic representation.[516] In any case, the 'great apes' in particular i.e. chimpanzees, gorillas and orang utangs have demonstrated remarkable abilities in using sign language and abstract symbols to *externally* represent an accurate grasp of actual and possible states of affairs.[517] Thus, various non-human animals exhibit a level of practical reason and autonomy which is appreciably similar to the conditions relevant for being granted a moral status approaching that accorded to prospective, purposive agents.

Acknowledgement of these relevant similarities in the autonomous agency of human and certain non-human animals generates constraints in deductive and explanatory coherence which rationally require adaptive moral agents to provide sufficiently for the well-being of such species. However, the scope of these adaptive learning capacities which humans share with their domestic animals should not be the primary factor justifying their status as recipients of moral concern. Rather than being contingent upon either their cognitive capacities, or their ability to be of some benefit to human interests, it is surely the capacity to experience pain and suffering which makes other sentient creatures worthy of moral consideration. In Bentham's famous words, "The question is not Can they *reason*? nor Can they *talk*? but, Can they *suffer*?"[518] Depending on the degree to which different species are capable of experiencing pain, the abovementioned criteria of deductive and explanatory coherence must also constrain the reasonability of rejecting moral principles which would seek to extend the same rights of protection to all beings "who are conscious and capable of feeling pain."[519] As Scanlon argues, "since human beings have good reason to avoid pain, they could reasonably reject principles that allowed others to inflict pain on them without good reason, or to fail to relieve their pain when they could easily do so." [520]

From this reasonable rejection of unnecessary pain being tolerated by members of our own species, it is incoherent to then deny that similarly painful experiences may nevertheless be tolerated in other species. As such, coherence constraints effectively *oblige* rational agents to acknowledge and apply the same prescriptive measures in the treatment of *all* needless pain, *regardless* of the recipient's ability to appreciate its own moral standing. Such reasoning is perfectly consistent with contractualist principles, as the reasonability which obliges human beings to respond to each others' sensitivity to pain, obliges them to respond with the same sensitivity to all creatures who are also potential *recipients* of some human action or inaction which causes them to suffer needless pain. Thus, while the rationality

and reasonability of human agents obliges them to include other sentient beings as recipients of moral concern, those beings who are incapable of grasping the concept of obligation obviously cannot be reciprocally obliged. However, it is precisely due to this deficiency in the power of their own rational and moral agency that we humans *owe* such creatures the same protective rights which we could not reasonably reject for ourselves. In this sense, "what we owe to each other" provides the same justification for what we owe to other creatures who, through no fault of their own, are bound to suffer *more* than *reason* would allow.

---

[458] i.e. meaning 'necessarily' or 'essentially', as opposed to 'inherently' or 'for its own sake'. See discussion at end of Chapter 4

[459] See Chapter 5, Section (i), pp.90.

[460] Haworth, op. cit., p.20

[461] Ibid.

[462] Rawls, *A Theory of Justice*, p.497

[463] Ibid., pp.503-504

[464] Ibid., p.268

[465] Rawls, *Political Liberalism*, pp.15-17

[466] Ibid., p.8

[467] John Rawls, *The Law of Peoples; with, The Idea of Public Reason Revisited* (Cambridge, Mass., Harvard University Press, 1999).

[468] Rawls, *A Theory of Justice*, p.504

[469] John Dewey, "Philosophy and Democracy," in *The Political Writings*, ed. D. Morris & I. Shapiro (Indianapolis; Cambridge, Hackett, 1993), p.46; also in *The Middle Works*, vol. 11.

[470] Rawls, *Political Liberalism*, pp.xlii-xlv

[471] Ibid., p.xliv

[472] Ibid., p.19

[473] Rawls, *A Theory of Justice*, pp.510-511

[474] Scanlon, op. cit., p.216

[475] Rawls, *Political Liberalism*, p.12

[476] Scanlon, op. cit., pp.213-218

[477] Rawls, *Political Liberalism*, pp.72-74

[478] Ibid., p.52

[479] Scanlon, op. cit., pp.191-193

[480] Ibid., p.193

[481] Ibid., p.211

[482] Ibdi., pp.211-212

[483] Ibid., p.211

[484] Gewirth, *Reason and Morality*, p.163

[485] Rawls, *Political Liberalism*, p.53

[486] Alan Gewirth, *The Community of Rights* (Chicago, University Of Chicago Press, 1996), pp.11-12

[487] Section (i) above

[488] DePaul, op. cit., p.162;

[489] Rawls, *Political Liberalism*, pp.17-18

[490] Ibid., pp.16-17

[491] Gewirth, *The Community of Rights*, p.76

[492] Ibid., p.77

[493] The *moral* significance of this heritage should not be taken to imply that one cannot reasonably reject clearly *immoral* causal factors. On the contrary, a child born as a result of the mother being raped, for example, has the best reason *to* reject both the irreparable harm already caused and any present and future factors which might encourage such acts.

[494] J. A. Hostetler, *Amish Society* (Baltimore, Johns Hopkins, 1980) p.298, cited in Sober & Wilson, op. cit., p.182

[495] Hardin, op. cit., pp.194-195

[496] Ibid., p.201; In *The Poverty of Historicism* (London, Routledge And Kegan Paul, 1961), Karl Popper warns that: "If the growth of reason is to continue and human rationality to survive, then the diversity of individuals and their opinions, aims and purposes must never be interfered with (except in extreme cases where political freedom is endangered). Even the emotionally satisfying appeal for a *common purpose*, however excellent, is an appeal to abandon all rival moral opinions and the cross-criticisms and arguments to which they give rise. It is an appeal to abandon rational thought."

[497] Erich H. Loewy, *Freedom and Community: The Ethics of Interdependence* (Albany, SUNY Press, 1993), p.42; It may be objected that such an extension of the notion of indebtedness defies the law of causation, as it is physically impossible for the acts or omissions of beings not yet in existence to affect the behaviour of those who presently exist. But this objection can only be sustained against beings who lack the physical and cognitive capacities to determine themselves as causal agents capable of foreseeing the likely consequences of their actions. Thus, rather than defying causal laws, it is *due to* the comprehension of such laws that human beings recognise their teleological agency through the power of inductive reasoning. Prospective, purposive agents are, by nature, beings whose actions are directed toward the intentional production of some specified future state. As long as there is sufficient evidence from past acts or omissions to indicate their likely effects upon a specifiable range of future conditions, it is disingenuous to claim that present generations owe nothing to their descendants. Such a claim rests upon a legalistic interpretation of indebtedness that conveniently ignores the indisputable truth noted above i.e. that present generations owe their existence to their ancestors' *future oriented* behaviour.

[498] Hardin, op. cit., pp.200-203

[499] Rawls, *Political Liberalism*, p.12

[500] Though, in many 'third world' countries, endemic conditions of poverty may provoke such a motivation

[501] Gewirth, *Reason and Morality*, p.122

[502] Ibid.

[503] See esp. Plato, op. cit., 460c

---

[504] See Edward Caudill, *Darwinian Myths: The Legends and Misuses of a Theory* (Knoxville, University Of Tennessee Press, 1997).

[505] Hume, *A Treatise on Human Nature*, II. I. XI. pp. 317-318

[506] Hume, *A Treatise on Human Nature*, II. I. XI. p.320

[507] Hume, *A Treatise on Human Nature*, II. II. XII. p.398

[508] Charles Darwin, *The Expression of the Emotions in Man and Animals*, (Chicago, University of Chicago Press, 1965), Ch. 2.

[509] Hume, *A Treatise on Human Nature*, II. II. XII. p.397

[510] Immanuel Kant, *Lectures on Ethics – Duties towards Animals and Other Spirits*, trans. L. Infeld (New York, Harper & Row, 1963), 240, cited in Mark H. Bernstein, *On Moral Considerability: An Essay on Who Morally Matters* (New York; Oxford, Oxford University Press, 1998), p.148

[511] Ibid p.148

[512] See De Waal, op. cit., pp. 93-94, on the social instincts of wolves which "explains their trainability for human purposes."

[513] Immanuel Kant, *The Critique of Judgment* (Amherst, New York; Prometheus Books) II, § 67, p.285

[514] Gewirth, *Reason and Morality*, p.122

[515] David Gauthier, "Morality, Rational Choice, and Semantic Representation: A Reply to My Critics," in *The New Social Contract: Essays on Gauthier*, ed. E. F. Paul et al (Oxford, Blackwell, 1988), p.173

[516] Reed, "The Intention to Use a Specific Affordance," p.57

[517] See *The Great Ape Project : Equality Beyond Humanity*, ed. P. Cavalieri & P. Singer (New York, St. Martin's Press, 1994), Chs. 4, 5 & 6.

[518] Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, ed. J. H. Burns & H. L. A. Hart (London; New York, Methuen, 1982), Ch.17, § 4, n. 1

[519] Scanlon, op. cit., p.179

[520] Ibid., p.181

CHAPTER NINE

ADAPTIVE SOCIAL CHOICE:
AN EVOLUTIONARY MODEL

In the previous chapter, as a result of acknowledging the evolutionary conditions which are necessary and sufficient for maintaining their adaptive capacities as prospective, purposive agents, it was shown that such agents could not reasonably reject moral principles which provide for the welfare of other non-human agents, whose common evolutionary heritage endows them with a similar capacity for experiencing pain and suffering. Likewise, human agents who are physically or mentally

incapacitated but still retain the capacity to suffer are all the more in need of assistance, on account of their limited ability to prevent or relieve their pain on their own behalf. As even the most adaptive individuals remain dependent upon the past, present and future choices and actions of other agents in conceiving, planning and realizing their goals, no autonomous agent can claim their level of autonomy and well-being to be functionally self-sufficient. Even Robinson Crusoe owes his survival prospects to the conceptual knowledge and practical skills learnt from his own society, while his prospects for being rescued are similarly dependent upon the navigational skills of his potential saviours who, in turn, are dependent upon continuing scientific improvements in the construction, equipment and power of ships etc. In the event that constant brooding upon his fate might lead him to develop such a thoroughly misanthropic perspective that he eventually comes to prefer his isolation, the fulfillment of that preference still depends on his ability to forestall or prevent the interference of others, which is itself subject to *their* ability to *refrain* from such interference.

In any case, while the eponymous Crusoe does indeed desire to be rescued, the absence of any social context might seem to completely erase any intelligibility in his claiming a right to be rescued. However, as Beyleveld argues on Gewirth's behalf, while "rights-claims presuppose a social context in directing what the claimant has a right to do, this social context constitutes only a subjunctive condition for the having of rights", so that in this example, "Robinson can say, "If Friday (or any other PPA) were to come along, then Friday (or any other PPA who came along) would have a duty not to interfere with my freedom and well-being.""[521] In justifying such duties, it is essential to recall Gewirth's distinction between mutuality and reciprocity.[522] As outlined in the previous chapter, for political purposes of mutual protection, the reciprocal benefits of many negative and positive rights and duties can be contractually justified in accordance with principles of rationality and reasonability. However, as PPAs agents also have *prior* rights to the basic level of autonomy and well-being needed for their purposes *independently* of those reciprocal relations. Applied to this case, from his own perspective Robinson must claim his agency rights "for the sufficient reason that he is a PPA, and so must grant these rights to Friday, because Friday is a PPA. From Friday's viewpoint we get similar reasoning."[523] Thus, it is their *mutual status as PPAs* that *logically* requires each to accord the same basic moral rights to the other, *independently* of the subsequent *political* process by which they are contractually obliged to adopt moral constraints.

As Dewey astutely observed, the self-restraint demanded by morality can only be achieved by an act of *self-assertion*, whereby the will effectively functions to inhibit the initiation of actions aimed at immediate self-gratification. This capacity for rational agents to adopt principles which override the impulses induced by the immediate prospect of pleasure is crucial to ensuring that the interests of all agents are protected from arbitrary interference, while also promoting habits of mutual aid which are not restricted to short-term, local benefits, or favouring family and friends to the exclusion of other agents' interests. Thus, it is the mutual selection of *rationally* considered preferences in wide equilibrium which induces adaptive agents to modify their depth and range of their moral concerns. However, as this process of intentional organic selection can only proceed with the aid of the indispensable functional capacities produced by natural selection, then the political expansion of moral consciousness must give full recognition to those functional constraints. Therefore, in the task of

selecting political principles which are most conducive to the development of adaptive moral agency, the capacity for pleasure and pain cannot serve as the sole criterion for determining the value of social policies or public goods.

The question of determining the most appropriate focus of external control in sustaining well-being is analogous to the issue of determining the level of selection primarily responsible for determining an organism's phenotypic success in a specific habitat.[524] Sen cites various theoretical approaches to assessing the agent's "social advantage", each focusing upon a distinct "informational base" e.g. personal utility focusing on pleasures, happiness, or desire fulfillment; absolute or relative wealth, focusing on income or commodities; negative freedoms focusing on power of non-interference; differences in the means of freedom focusing on 'primary goods', or in resources focusing on justice defined by their equal distribution.[525] As Sen suggests, these sources of information in gauging social advantage are unnecessarily restrictive in limiting the potential variety of objects in the "evaluative space" available to individual agents.[526] For this reason, Gewirth can be criticised for subordinating well-being to the agent's choices, so that a person with few practical options would be obliged to accept conditions which, in effect, *minimise* their prospective freedom and well-being. Endorsing Sen's point in this regard, Held notes that a woman's dutiful decision to stay in an arranged marriage, in spite of the continuing misery it brings her, can scarcely be said to have expressed a *valued preference*.[527]

This failure to consider the evaluative import and social context of the PPA's judgement that '*E* is good' has already been highlighted and addressed, necessitating a significant amendment to the pragmatic conditions of agency, wherein the value of autonomy is given precedence over freedom of choice. In addition to its function as a necessary condition for adaptive agency, the autonomous selection and fulfillment of valued ends leads to an increase in the level of well-being which, in turn, creates further opportunities for purpose fulfillment, yielding still further increases in autonomy and well-being. However, as this synergistic enhancement of adaptive agency is subject to ecological and epistemic constraints, as well as incompatibilities among competing purposes, these evolving contingencies must also be factored into any realistic measure of an individual's 'social advantage', insofar as they give rise to *disadvantages* in the capacity to function adaptively. This is precisely where Sen's focus on the functional *capabilities* of agents show *its* adaptive fitness as the most appropriate scheme for determining the degree to which any particular individual's level of well-being is sufficient to enable them to consistently achieve their preferred ends.

Sen assesses the relative contributions of freedom, well-being and agency itself in the fulfillment of different purposes, paying special attention to the most basic constituents of successful action: freedom and achievement. Connecting these elements to well-being and agency yields four distinct categories for both intrapersonal and interpersonal comparisons of functional capability i.e. well-being achievement, well-being freedom, agency achievement and agency freedom.[528] While Sen allows the practical meaning of freedom to include the absence of constraints, as in the capability afforded by a "malaria-free life", or "freedom from hunger", it is the scope of the individual's freedom to choose which is crucial to enhancing their well-being and agency.[529] Under Sen's scheme, the autonomy of adaptive agents would therefore be perfectly accommodated under the description of 'agency freedom', given the increased capacity for self-control that such freedom brings. More

importantly, the dynamic interdependence among these four capabilities tends to produce the same synergistic effects and practical constraints manifested in the evolution of adaptive agency. Thus, Sen notes that:[530]

> e.g. an increase in well-being, other things being equal, will involve a higher agency achievement…In addition, a failure to achieve one's *non*-well-being objectives may also cause frustration, thereby reducing one's well-being…Similarly, more freedom (either to have well-being or to achieve one's agency goals) may lead one to end up achieving more (respectively, of well-being or of agency success), but it is also possible for freedom to go up while achievement goes down, and vice versa.

Evidently, the process of these fluctuations in the agent's functional capabilities closely resembles the mutual selection process whereby different levels of autonomy and well-being constrain the agent's own selection of practical purposes which, in turn, reduce the opportunities for enhancing autonomy and well-being. However, it is particularly important to note the implications of Sen's recognition that either freedom or achievement may be advanced at the other's expense. Having already argued against Gewirth that valuing freedom *simpliciter* does not rule out the achievement of self-destructive or immoral ends, the same criticism would arise here if Sen were to cite the agent's freedom to choose as a sufficient condition for the achievement of moral agency. But in distinguishing such freedom from that which is directed toward achieving one's *agency goals*, he describes such goals as those "that a person has reasons to adopt, which can *inter alia* include goals other than the advancement of his or her own well-being."[531] Furthermore, in noting that "the effect of 'other-regarding concerns on one's well-being has to operate *through* some feature of the person's well-being", Sen shows his understanding of the need to interpret the agent's moral disposition in terms of its consequences for that individual's prospective well-being.[532]

In revealing how the agent's moral interests are essentially realized in and through their own personal capabilities and functional achievements in a specific environment, Sen's methodology succeeds in preserving the inherently *social* dimensions of rational choice. Given that the standard theories of rational choice proceed from an abstract conception of agents as variously atomistic, egoistic, or mutually disinterested, it is not surprising that such perspectives have great difficulty in accounting for the empirical fact that the majority of human agents do *not* have to be coerced into accepting and fulfilling positive obligations to those from whom they need not necessarily benefit in return. However, this criticism should not be assumed to imply that the development of moral agency cannot be defended on primarily rational grounds. On the contrary, as the reasonability which facilitates the acceptance of moral principles has already been shown to be derived from and constrained by principles of rational coherence, the emergence of a disposition toward moral thought and action is dependent upon the agent's adequate grasp of rational principles. To be sure, even a thorough appreciation of the prudential benefits afforded by the application of such principles may not be sufficient to elicit the degree of sympathy or moral 'sentiment' required by the standard Humean account of motivation. But there are alternative strategic models of rational choice which do not rest

upon the unfounded assumption that an initial deficiency in the strength of emotional identification should necessarily impede the active development of moral agency.

As Gauthier's rational agents are mutually unconcerned or "non-tuistic" utility maximisers, their acceptance of moral constraints upon their considered preferences is motivated by a rational recognition of the superior benefits afforded to each by adopting a policy of "constrained maximisation". In contrast to the standard "straightforward" maximiser of rational choice theory, who is "disposed to maximise her satisfaction, or fulfil her interest, in the particular choices she makes", a constrained maximiser is "disposed to comply with mutually advantageous moral constraints, provided he expects similar compliance from others."[533] Gauthier insists that his depiction of human beings as a fundamentally asocial species - *'Homo Economicus'* – is not, of course, intended as an accurate representation of human beings, but is only assumed for the provisional purpose of ensuring that "morality is not affectively dependent, so that it speaks directly to reason and not to particular, contingent emotions or feelings."[534] In this respect, Gauthier simply assumes the unqualified validity of Kant's contention that "morality binds independently of the nature and content of our affections", so that "one is not then able to escape morality by professing a lack of moral feeling or concern, or some other particular interest or attitude."[535] Thus, in seeking to demonstrate that strategic rationality *alone* may suffice to generate a binding commitment to the constraints of morality, Gauthier must explicitly avoid any *theoretical* appeal to sentiments or benevolent attitudes in motivating that commitment. However, this does not imply that *in practice* agents' actual affective capacities have no motivational role in sustaining their moral dispositions. Therefore, even while endorsing Kant's dismissal of any psychological interest that might be invoked in the rational justification of binding moral duties, Gauthier does subsequently allow Humean sentiments a role in eliciting an agent's willing performance of those duties. In this regard, Gauthier distinguishes between agents whose "affective capacity for morality" allows them to endorse constraints to which they are already predisposed and those with a "capacity for affective morality" who accept such constraints only because of their instrumental value in furthering such agents' own non-tuistic concerns.[536]

Gauthier has been duly criticised for failing to give an adequate account of the positive motivational force of the emotions in guiding moral responses and judgements.[537] Philosophers have typically been suspicious of the evident malleability of human emotions in guiding judgement, and an evolutionary analysis of their serviceability in facilitating co-operative commitments certainly gives strong empirical support to this intuition. There is no shortage of historical and experimental evidence demonstrating the most appalling excesses to which human beings may be driven, when the capacity for mutual empathy functions only to reinforce prejudices and hostilities of one group against another.[538] Likewise, emotional ties between individuals may work to maintain injustices by the threat of guilt and shame in transgressing cultural norms, as in Held's example of the woman who *feels* duty bound to stay in an oppressive marriage. Gauthier himself cites the exploitation of women in this regard, as instantiating a further reason to avoid seeking moral guidance by appealing to agents' affective capacities.[539] However, Gauthier cannot be excused for his own uncritical appeal to Kant's assertion that such affective capacities or interests must be thoroughly ignored, even where they are perfectly adapted to ensuring the reliable performance of the moral duties which reason would require.

Responding to Gauthier's focus upon the negative influence of affectivity, Laurence Thomas issues a reminder of the indispensable evolutionary value of human beings' "considerable capacity for affection", without which "life as we know it simply could not be. Romantic loves, friendships, and the parent-child relationship would be radically different in the absence of the capacity for affection." In support of these remarks, Thomas cites Maynard Smith's 'hawks and doves' game-theory scenario as evidence that the competition and conflict in a population of selfish individuals would cause a significant decline in their numbers, thereby fostering conditions conducive to the flourishing of altruists[540] (notwithstanding the converse tendency in such population dynamics which also ensures that the numbers of doves is bound to be restricted by the presence of a proportion of hawks). Yet Gauthier, like Rawls, also refers to the evidence from Trivers' theory of reciprocal altruism in giving explanatory coherence to his own argument that "a population of constrained maximisers would be rationally stable; no one would have reason to dispose herself to straightforward maximization." Strong analogical coherence is also established in Gauthier's conception of constrained and straightforward maximisation as "parallel to genetic tendencies to reciprocal altruism and egoism."[541] Nevertheless, his claim that the rational interests of SMs is sufficient to effect their permanent conversion to CMs is still liable to be rendered incoherent, due to the lack of an affective influence in lending *psychological* stability to their dispositions. Given that SMs only possess a *capacity* for affective morality, their newly acquired disposition for constrained maximisation is crucially dependent upon that capacity being sufficiently well developed and appropriately directed. This is, of course, the basis of Sen's argument that human capabilities are liable to remain undeveloped, unused or misdirected, if no attention is given to their successful functioning in the agent's actual environment. Therefore, the mere possession of an affective capacity is scarcely sufficient to ensure that the agent regularly adheres to moral constraints when and where required.

It is perhaps in appreciation of this weakness that Gauthier again appeals to Trivers' argument that "natural selection will favour the development of guilt, as a device motivating those who fail to reciprocate to change their ways in future" and that "we may expect that in the process of socialization, efforts will be made to develop and cultivate each person's feelings so that, should she behave as an SM, she will experience guilt."[542] This latter admission is particularly illuminating, both for and against the coherence of Gauthier's model. While it certainly undermines the Kantian justification for his initial representation of ideally rational agents as asocial, non-tuistic SMs, it does provide an accurate empirical account of the normal process of moral development which is not only consistent with Gauthier's contractual reconstruction of morality as the internalisation of normative constraints, but also delivers substantial explanatory coherence from the acknowledged connection to the evolutionary conditions favouring those constraints.

The transition from SM to CM can thus be understood as analogous to the ontogenetic process of moral development described by Piaget and elaborated in Kohlberg's stages, wherein the child's innately self-interested motive in being rewarded for morally acceptable behaviour is eventually reassessed and assimilated in maturity in terms of its *reasonability* as a scheme for the *mutual* benefit of all members of society. But the active functioning of empathic emotions is therefore of vital importance in facilitating a commitment to moral actions which might otherwise be liable to severe or

prolonged lapses, particularly if the agent's adoption of a moral disposition is still ultimately valued for the superior benefits it affords. To that end, emotions themselves can be rationally evaluated in terms of their coherence with moral principles and goals. However, moral constraints upon the emotions must be consistent with the evolved sensitivities to pain and suffering, in order to avoid the imposition of demands which are inimical to the basic needs and desires of all agents. For instance, the reasonability of policies which restore equality in agents' capabilities to successfully nurture and educate their children cannot be made to cohere with policies requiring parents to suppress the innate emotions which strongly induce and *support* that concern. However, the reasonability in policies requiring equality of opportunity is thoroughly inconsistent with the encouragement of such emotional attachments, which can be discounted without harm to those who would otherwise prefer that nepotism be permitted to advance their careers or educational opportunities. More importantly, the impartiality which constrains the operation of such emotions in public policies *gains* emotional coherence from the mutual trust which is enhanced as a result.

From an evolutionary understanding of their action-guiding functions, the traditional division between reason and emotion cannot be sustained. In spite of his defence of this distinction in Kantian terms, it is ironic that Gauthier (no less than Kant himself) should have neglected the applicability of another *Kantian* axiom of practical reason that is incontestable – 'ought implies can'. Neuroscientific and psychological evidence show that when the areas of the brain controlling higher cognitive processes are deprived of their connections to the emotion centres such as the amygdala, decision-making is seriously impaired.[543] As Thagard notes, even if it were possible, to "turn off your amygdala…to do so would cut your analytical decision making off from crucial emotional information about what really matters to you."[544] Given the functional capabilities afforded by these biological systems, Gauthier's rational choice conditions can only benefit from allowing agents to be attentive to the emotional significance of their prospective actions, at least insofar as those emotions are able to lend motivational coherence to the reasonability of moral commitment.

The adoption of a moral disposition for constrained maximisation is therefore a political strategy which cannot reasonably be rejected by prospective, purposive agents, having acknowledged the need to protect and advance their potential capabilities to function as adaptive moral agents. However, as the adaptive benefits to each individual's scope for purpose fulfillment can only be guaranteed on the condition that others can be trusted to be equally disposed to abide by such constraints, then commitment to such a strategy is only reasonable to the extent that such trust is well founded. Thus, by insisting that rational agents are primarily motivated to accept moral constraints only insofar as it advances their own interests, Gauthier's account is not sufficiently robust to prevent committed moral agents from being disadvantaged and exploited by those who choose to ignore such obligations when it is their rational self-interest to do so. Alternatively, SMs may deprive CMs of valuable opportunities and resources through deception, by strategically simulating the CM's moral behaviour to gain the latter's trust. As illustrated in the scenario where the self-styled demi-god Baal escapes to a life of luxury in the Bahamas at the expense of his ascetic but gullible disciple Zeke, adherence to self-imposed moral demands can become a liability for a moral agent's prospects.

As Gauthier notes in reference to Trivers' argument: "…natural selection will favour the development of the capacity to detect merely simulated altruism. This, of course, corresponds to our claim that constrained maximizers, to be successful, must be able to detect straightforward maximizers whose offers to co-operation are insincere."[545] While many experiments have indeed verified the inheritance of such "cheater detector" mechanisms, their efficiency is limited to identifying transgressors in small, tribal populations or social domains where repeated interactions can be readily monitored.[546] If the agent's local environment does not permit access to information which might facilitate the detection and public warning of such threats, or the stability of a hierarchical social structure is itself dependent upon inducing competition which rewards such parasitism, then moral agents must resort to costly punitive measures to minimise their exposure to such risks. However, all such measures are still premissed upon the false Hobbesian depiction of persons who have never evolved any affective mutual interests other than the need to defend themselves from each other's avarice.

The widespread emergence and stability of reciprocal altruism would not be possible if mature human agents were only capable of adhering to moral constraints when it is to their maximal advantage. Over each person's life span, their adaptive capacities and prospective purposes cannot be maintained and achieved by choosing their options *parametrically*. McClennen summarises the conditions of parametric reasoning as follows:[547]

(1) the agent is presumed to have an antecedently specifiable preference ordering over the set of all possible outcomes of action

(2) rational choice consists in selecting a feasible course of action whose associated outcome is maximally preferred and

(3) the set of background considerations that condition any moment of choice function essentially only to restrict the set of feasible actions – they do not shape in any way the agent's preference ordering over actions

As McClennen argues, this traditional conception of rational choice is fundamentally incompatible with the "notion that the agent is a being who continues over time, with concerns that have some continuity to them…What is characteristic of such agents is that his *ex post* preferences among available actions are disciplined or shaped by what he judges, from the perspective of plans taken as wholes, to be the best plan to pursue."[548] This dynamic conception of the adaptive agent's capacity to intelligently constrain their future choices on the basis of an initial set of preferred outcomes has already been highlighted as an effective strategy for combating the evolved bias toward reasoning myopically, in contexts which favour choosing immediately achievable rewards over more highly valued long-term goals. With regard to freedom as a necessary law, Dewey notes that:

When we use the law to foresee consequences and to consider how they may be averted or secured, then freedom begins. Employing knowledge of law to enforce desire in execution gives power to the engineer. Employing knowledge of law in order to submit to it without

further action constitutes fatalism, no matter how it be dressed up….Human desire and ability cooperate with this or that natural force according as this or that eventuality is judged better. We do not use the present to control the future. We use the foresight of the future to refine and expand present activity. In this use of desire, deliberation and choice, freedom is actualized."[549]

With regard to freedom as a necessary condition for securing the autonomy of adaptive moral agency, McClennen's 'Principle of Dynamic Consistency' thus succeeds as an accurate model of the PPA's deliberative task in controlling the preferential selection of means and ends needed to promote the future well-being, while also serving as a means of allowing agents to trustfully co-ordinate their purposive activities, by the mutual adoption and cultivation of a *resolute* disposition towards the fulfillment of those general moral principles which cannot fail to pass the contractual reasonability test.

McClennen's strategy of "resolute choice" is necessarily "context sensitive" in that each agent's adaptive prospects can only be measured from within the context of the moral constraints that have already structured the agent's future options, as long as these remain feasible when the time arrives to act upon them.[550] Accordingly, when the fluctuations in the capability functionings of agents in different ecologies are evaluated and compared in terms of their *dynamic consistency* in promoting a stable and sustainable level of purpose fulfillment, then persistent inequalities in agents' adaptive capabilities can be prevented from evolving without compromising the agent's autonomy of choice. In this respect, it is important to distinguish the strategy of resolute choice from that exemplified by Ulysses and the Sirens. In such circumstances where an akratic change in future preferences can be reliably anticipated, those most considered preferences can be effectively secured in advance, by employing a precommitment strategy that imposes *occurrent* and *local* restrictions on the agent's capabilities, in order to preserve and enhance their *prospective* and *global* functionings. However, as this self-restraint in accessing ecological affordances may be detrimental to a person's current level of well-being and its enforcement often involves elaborate preparations and inducements which may not be procurable in impoverished circumstances, such measures can only be used to advance the adaptive prospects of those who already have access to a surplus of convertible goods. By contrast, the agent who makes a resolute choice need not have recourse to such exogenous constraints, as their future preferences are *endogenously* adapted to those contexts where a co-operative disposition is more advantageous to the totality of the agent's functionings.[551] However, it must be remembered that the acceptance of both internal and external constraints is directed toward the goal of ensuring the level of *adaptive* autonomy and well-being needed to function as a *reflective* moral agent. Therefore, if such constraints are to be intelligently directed toward moral goals, the reasonability of resolute choice necessarily excludes the adoption of a set of absolute, categorical imperatives which are insensitive to contingencies in person and situation. It is in this respect that Dewey notes: "The ethical import of the doctrine of evolution is enormous" as it involves "continuity of change….Significant stages in change are found not in access of fixity of attainment but in those crises in which a seeming fixity of habits gives way to a release of capacities that have not previously functioned: in times that is of readjustment and redirection."[552]

In Sen's categories, "a person's *capability set* can be defined as the set of functioning vectors within his or her reach."[553] These combinations of "doings and beings" range from the most basic functional necessities, such as good health and nutrition, to more elusive social commodities such as the achievement of self-respect.[554] As noted above, the superiority of Sen's analysis consists in its focus upon both freedom *and* achievement, thus enabling the comparative assessment of agents' overall capabilities to function successfully in a specific environment. Of the four functional vectors identified by Sen, *well-being freedom* is of primary importance, as Gewirth has also noted in terms of their material necessity for the successful realization of any purposive action. However, by observing the effects of their mutual interaction upon the development of personal agency, Sen avoids Gewirth's tendency to assess their contributions independently of such evolving effects. As an example of this freedom to achieve well-being, Sen considers a case where two persons are both starving to an extent of being equally deprived of nourishment. But if *A*'s starvation occurs as a result of her inability to obtain enough food, while that of a wealthy man *B* is induced by the religious practice of fasting, then the *freedom* to choose this occurrent reduction in well-being is denied in *A*'s case, thus reflecting an inequality of capability functioning which would not be observable if well-being is assessed in terms of an index of single measure functionings such as the use of primary goods.[555]

In this case, *B*'s freedom to *choose* fasting is indeed a situation where being resolutely predisposed to complete that task is beneficial from *B*'s global perspective of his autonomy and well-being, as defined by his conception of the good. As long as his fasting cannot be foreseen to adversely affect the autonomy and well-being of other agents, his resolute choice to adhere to such a regime allows him to achieve dynamic consistency among his considered preferences, thus enabling greater adaptive control in determining his own destiny and self-fulfillment. However, the scenario may be such that, despite his occasional acts of fasting, a substantial proportion of *B*'s considerable wealth is obtained by paying a pittance in wages to those such as *A*, who is employed in the local factory which *B* owns and manages. In that case, the "personal spheres" in which *A* and *B* operate are causally interactive to a degree where the well-being freedom and agency of each is asymmetrically enforced by the effective dominance of *B*'s strategy of paying 'starvation' wages to *A*.

The dynamics of such asymmetrical relations are not unlike those described in Hegel's analysis of the dialectical fluctuations in master/slave relations. Although these evolutionary dynamics operate through naturalistic organic selection pressures which cannot be reconciled with the idealism of Hegel's dialectic, both interpretations reveal the dominant party's *dependence* on those they exploit. As a result of this dependence, even though the dominant party's freedom and well-being is advanced by largely depriving the other party of the capability to use their own share of such essential goods for their own purposes, the dynamic moral effects of such exploitation are not limited to those incurred by the exploited party. This is neatly illustrated in another of Sen's examples in which person *B* drowns as a result of *A*'s failure to rescue him, despite *A*'s proximity. In this case, as it is evident that *B* cannot swim, while *A* is a champion swimmer, in addition to the suffering and death of *B* and its effects upon his loved ones, *A*'s failing reveals a critical loss of functioning in his moral agency.[556]

In such contexts, the adoption and cultivation of a resolute moral disposition can succeed in preventing such lapses in moral agency, even at the cost of a temporary loss in the agent's occurrent

level of well-being. In this regard, Sen considers a more positive variation on the aforementioned case: you are sitting on a riverbank, eating a sandwich, when you see a drowning man. As a capable moral agent, you will not hesitate to "chuck your sandwich, jump into the cold river and haul the man out."[557] As Sen notes, although your actions cause an occurrent reduction in your well-being, in suffering from the cold and being prevented from enjoying your lunch, these losses are immeasurably outweighed by a significant expansion in your agency freedom.[558] This example provides an instructive contrast to that of the donut-muncher who fails to develop a global and dynamic perspective upon the formation of his preferences, despite his avowed resolution to make a donation to famine relief.

In conclusion, the organic cultural conditions needed to sustain the development of adaptive moral agency can be summarised in three steps, the first two of which are analogous to those in which the Baldwin Effect is manifested in the evolution of phenotypic plasticity:

(1)   Phenotypic Plasticity:

Innate behavioural flexibility (minimal autonomy) in applying heuristics to recognise and adaptively respond to changeable environmental hazards and esp. to locate and access common *affordances* (i.e. resources, opportunities for enhancing autonomy & well-being), leads to the evolution of social norms, conventions e.g. moral & political principles, virtues etc. Moral consciousness & autonomy in population thereby raised to level of reflection adapted to viability of local affordances.

(2)   Ecological Modification:

Frequency and stability of moral and political conventions (Social Contract) alters the pragmatic affordances of the social environment, to the effect that it becomes progressively more 'user-friendly' for subsequent generations of individuals to develop dispositions for commitment to such conventions.

(3)   Internalisation of Rules/Norms:

Generations of strong selection pressures favouring dispositions for efficient adoption of inherited norms and conventional practices promotes a tendency to uncritically absorb and reproduce them, leading to the effective *loss* in the optimal power of reflective autonomy and level of well-being achieved in stage (2)

As human agents' advanced level of adaptive autonomy and well-being is sustained only by avoiding critical lapses into the heteronomous behaviour in (3), the social contract in (2) must be rationally remodelled to promote an *ecological* and *dynamic* perspective of moral and political agency sufficient to ensure that internalised dispositions in (3) are autonomously cultivated and intelligently directed.

As noted in Chapter One,[559] Godfrey-Smith duly highlights the crucial importance of avoiding the final step whereby natural selection processes eventually lead to the evolution of genotypes which effectively "canalize" behavioral responses, leading to a loss of flexibility. Just as "genetic assimilation is the enemy of flexibility"[560] in the evolution of agency for organisms inhabiting complex, changing environments, the inflexible "internalisation" of moral norms or rules is the enemy of adaptive moral agency. In any case, the analogy with genetic assimilation is not in fact necessary, as Baldwin himself "placed greatest emphasis on the survival benefits of psycho-genetic traits: ontogenetic plasticity gained through trial-and-error learning, imitation, instruction, and conscious intelligence."[561]

The method of wide reflective equilibrium which guides this internally global and dynamic perspective is reflected outwardly in the moral agent's similar perspective on their adaptive capabilities in a global ecology. Accordingly, the task of restoring equality in the minimal conditions required for the optimal fulfillment of such capabilities can only be secured through the political protection of global laws and institutions which must evolve in response to changing conditions. Various indices of rights have been formulated and revised in order to distinguish between capabilities on the basis of their degree of urgency for meeting agent's universal needs. From an evolutionary perspective, the exclusive focus upon fixed needs or ultimate ends fails to account for the dynamic constraints and ecological variations to which all agents must adaptively respond.[562] Yet, with these factors in mind, even imperfectly rational agents could not reasonably reject the need to expand their moral horizons. For it is through the fulfillment of *this* need that all other needs can be fulfilled.

---

[521] Beyleveld, *The Dialectical Necessity of Morality*, p.161

[522] See Ch. 8, Section (ii), p.154, note 34

[523] Beyleveld, *The Dialectical Necessity of Morality*, p.465

[524] See Williams, *Natural Selection*, Ch.4; Sober & Wilson, op. cit., Ch.3

[525] Amartya Sen, "Capability and Well-being," in *The Quality of Life*, p.30

[526] Ibid., pp.32-33

[527] Virginia Held, "The Normative Import of Action," in *Gewirth*, pp.17-18

[528] Sen, "Capability and Well-being," p.49

[529] Ibid., p.45

[530] Ibid., pp.35-36

[531] Ibid., p.35

[532] Ibid., p.36

[533] Gauthier, *Morals by Agreement*, p.15

[534] Ibid., p.329

[535] Ibid., p.103

[536] Ibid., pp.326-328

[537] See esp. essays by Christopher Morris: "The Relation between Self-Interest and Justice in Contractarian Ethics," pp.119-153 and Laurence Thomas: "Rationality and Affectivity," pp.154-172 in *The New Social Contract*.

[538] See Hardin, op. cit., Ch.6

[539] Gauthier, *Morals by Agreement*, p.11

[540] Laurence Thomas, op.cit., in *The New Social Contract*, p.163

[541] Gauthier, *Morals by Agreement*, p.187

[542] Ibid., pp.187-188; See David Copp, "Contractarianism and Moral Skepticism," in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, p.222 n.39, where Gauthier now disclaims the need to characterise agents as non-tuistic. But cf. Howard Kahane, *Contract Ethics: Evolutionary Biology and the Moral Sentiments* (Lanham, Md.; London, Rowman & Littlefield, 1995), p.122. In reference to the scenario of the Woody Allen film, *Crimes and Misdemeanors*, "where one of the characters has his former lover murdered when she threatens to expose their affair to his wife", Gauthier reportedly remarked that his theory would condone such an action, requiring the character to "bite the bullet."

[543] Thagard, op. cit., pp.212-213

[544] Ibid., p.214

[545] Ibid., p.187

[546] See Gigerenzer, op. cit., pp.213ff. & 229-230

[547] Edward F. McClennen, "Constrained Maximization and Resolute Choice," in *The New Social Contract*, p.105

[548] Ibid., pp.111-112

[549] Dewey, Human *Nature and Conduct*, pp.312-313.

[550] McClennen, "Constrained Maximization and Resolute Choice," p.110

[551] Ibid., p.111

[552] Dewey, *Human Nature and Conduct*, p.284

[553] Sen, "Well-being, Agency and Freedom: The Dewey Lectures 1984," in *Journal of Philosophy* vol. 82, p.201

[554] Sen, "Capability and Well-being," p.31

[555] Sen, "Well-being, Agency and Freedom," pp.201-202

[556] Ibid., p.220

[557] Ibid., p.206

[558] Ibid., p.207

[559] See note 11.

[560] Godfrey-Smith, "Between Baldwin Skepticism and Baldwin Boosterism," p.58

[561] Celia Moore, "Development and Individual Acquisition of Traits," in *Evolution and Learning*, p.116

[562] The fact that different conceptions in measuring well-being have themselves evolved in response to deficiencies only testifies to the merits of Dewey's experimental approach. See Paul Streeten, "Shifting Fashions in Development Dialogue," pp.68-81 and Selim Jahan, "The Evolution of the Human Development Index", pp.128-139 in *Readings In Human Development: Concepts, Measures and Policies for a Development Paradigm*, ed. S. Fukuda-Parr and A. K. Shiva Kumar, New Delhi, Oxford University Press, 2003.

BIBLIOGRAPHY

Alexander, R. D. *The Biology of Moral Systems*. New York, Aldine De Gruyter, 1987.

Argyle, M. *Cooperation : The Basis of Sociability*. London, Routledge, 1991.

Aristotle. *The Complete Works of Aristotle*. Revised Oxford Translation, ed. J. Barnes. Princeton, Princeton University Press, 1984.

Aristotle. *The Nicomachean Ethics*, trans. H. Rackham. London, Wordsworth, 1996.

Axelrod, R. M. *The Evolution of Cooperation*. New York, Basic Books, 1984.

Axelrod, R. M. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton, Princeton University Press, 1997.

Batson, C. Daniel. *The Altruism Question: Toward a Social-Psychological Answer*. Hillsdale, NJ; Lawrence Earlbaum Associates, 1991.

Bentham, J. *An Introduction to the Principles of Morals and Legislation*, ed. J. H. Burns and H. L. A. Hart. London; New York, Methuen, 1982.

Bernstein, M. H. *On Moral Considerability: An Essay on Who Morally Matters*. Oxford, Oxford University Press, 1998.

Bevan, D. (ed.) *Modern Myths*. Amsterdam & Atlanta GA; Rodopi, 1993.

Beyleveld, D. *The Dialectical Necessity of Morality: An Analysis and Defense of Alan Gewirth's Argument to the Principle of Generic Consistency*. Chicago, University of Chicago Press, 1991.

Blackmore, S. J. *The Meme Machine*. Oxford, Oxford University Press, 1999.

Bloch, M. L. B. *Feudal Society*. 2nd ed., trans. L. A. Manyon. London, Routledge & Kegan Paul, 1962.

Boehm, C. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Cambridge, MA; London, Harvard University Press, 1999.

Bogdan, R. J. *Interpreting Minds: The Evolution of a Practice*. Cambridge, Mass., MIT Press, 1997.

Boyden, S. *Western Civilization in Biological Perspective: Patterns in Biohistory*. Oxford, Clarendon Press, 1987.

Boylan, M. (ed.) *Gewirth: Critical Essays on Action, Rationality, and Community*. Lanham, MD, Rowman & Littlefield, 1999.

Bratman, M. *Intention, Plans, and Practical Reason*. Cambridge, Mass., Harvard University Press, 1987.

Camus, A. *The Myth of Sisyphus*. Harmondsworth, Penguin Books, 1975.

Caudill, E. *Darwinian Myths: The Legends and Misuses of a Theory*. Knoxville, University of Tennessee Press, 1997.

Cavalieri, P. and Singer, P. (eds.) *The Great Ape Project: Equality Beyond Humanity*. New York, St. Martin's Press, 1994.

Cohen-Solal, A. *Sartre: A Life*. London, Heinemann, 1987.

Coveney, P. and Highfield, J. R. L. *Frontiers of Complexity: The Search for Order in a Chaotic World*. London, Faber, 1996.

Daniels, N. (ed.) *Reading Rawls: Critical Studies on Rawls' A Theory of Justice*. New York, Basic Books, 1975.

Danielson, P. (ed.) *Modeling Rationality, Morality, and Evolution*. Oxford, Oxford University Press, 1998.

Darwall, S. *The British Moralists and the Internal 'Ought': 1640-1740*. Cambridge, Cambridge University Press, 1995.

Darwin, Charles. *The Descent of Man*; *and Selection in Relation to Sex*. 2nd ed. Amherst, NY; Prometheus Books, 1998; original ed. New York, Cromwell, 1874.

Darwin, Charles. *The Expression of the Emotions in Man and Animals*. Chicago, University of Chicago Press, 1965. First published, London, J. Murray, 1872

Darwin, Charles. *On the Origin of Species by Means of Natural Selection or, The Preservation of Favoured Races in the Struggle for Life*, 1st ed. Ware, Wordsworth Classics ed., 1998. Original ed. 1859.

Dawkins, R. *The Selfish Gene*. London, Paladin Granada Publishing Ltd., 1976.

Dawkins, R. *The Extended Phenotype: The Gene as the Unit of Selection*. Oxford, W.H. Freeman, 1982.

Deacon, T. W. *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York, Norton, 1997.

Dennett, D. C. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. London New York, Simon & Schuster, 1995.

DePaul, M. R. *Balance and Refinement: Beyond Coherence Methods of Moral Inquiry*. London, Routledge, 1993.

De Waal, F. *Chimpanzee Politics: Power and Sex among Apes*. Baltimore & London, Johns Hopkins University Press, 1989.

De Waal, F. *Good Natured: The Origins of Right And Wrong in Humans and Other Animals*. Cambridge, Mass; London, Harvard University Press, 1996.

Dewey, J. *Art as Experience*. New York, Capricorn Books, 1958.

Dewey, J. *The Essential Dewey*, vols. 1 & 2, ed. L. Hickman and T. M. Alexander, Bloomington, Indiana University Press, 1998

Dewey, J. *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. Boston; New York, Heath, 1933.

Dewey, J. *Human Nature and Conduct.* New York, Henry Holt, 1922.

Dewey, J. *The Middle Works, 1882-1898*, vol. 2, ed. J. Boydston, Carbondale, Southern Illinios University Press, 1976-1983.

Dewey, J. *The Moral Writings of John Dewey* ed. J. Gouinlock, New York, Macmillan, 1976.

Dewey, J. *The Political Writings*, ed. D. Morris and I. Shapiro. Indianapolis; Cambridge, Hackett, 1993.

Dewey, J. *The Quest for Certainty: A Study of the Relation of Knowledge and Action*. New York, Capricorn Books, 1960.

Dewey, J. *Theory of Valuation*. Chicago, The University of Chicago Press, 1939.

Douglas, K. "Born to Trade." *New Scientist*, Vol. 183, No. 2465, 18 September 2004.

Durham, W. H. *Coevolution: Genes, Culture, and Human Diversity*. Stanford, Stanford University Press, 1991.

Erasmus, D. *Adages*, trans. M. M. Phillips, Toronto, University of Toronto Press, 1974.

Foot, P. (ed.) *Theories of Ethics*. Oxford, Oxford University Press, 1967.

Foot, P. *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford, Basil Blackwell, 1978.

Frankfurt, H. G. *The Importance of What We Care About: Philosophical Essays*. Cambridge, Cambridge University Press, 1988.

Fried, M. H. *The Evolution of Political Society*: *An Essay in Political Anthropology*. New York, Random House, 1967.

Fukuda-Parr, S. and Shiva Kumar, A. K. (eds.) *Readings In Human Development: Concepts, Measures and Policies for a Development Paradigm*. New Delhi, Oxford University Press, 2003.

Gauthier, D. P. *Morals by Agreement*. Oxford, Clarendon Press, 1986.

Gewirth, A. *The Community of Rights*. Chicago, University of Chicago Press, 1996.

Gewirth, A. *Human Rights: Essays on Justification and Applications*. University of Chicago Press, Chicago and London, 1982.

Gewirth, A. "Is Cultural Pluralism Relevant to Moral Knowledge?" *Social Philosophy and Policy* 11, no.1, 1994. 22-43

Gewirth, A. "The Problem of Specificity." *Biology and Philosophy*, 1, 1986, 297-305.

Gewirth, A. *Reason and Morality*. Chicago, University of Chicago Press, 1978.

Gewirth, A. "Rights and Virtues." *Review of Metaphysics* 38 1985, 739-762

Gibbard, A. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, Mass., Harvard University Press, 1990.

Gigerenzer, G. *Adaptive Thinking: Rationality in the Real World*. Oxford, Oxford University Press, 2000.

Ginger, A, Hobbs J. and Lewis H. (eds.) *Selected Interdisciplinary Essays on the Representation of the Don Juan Archetype in Myth and Culture*. Lewiston; Queenston; Lampeter; Edwin Mellen Press, 2000.

Glover, J. *Humanity: A Moral History of the Twentieth Century*. New Haven, Yale University Press, 2000.

Godfrey-Smith, Peter. *Complexity and the Function of Mind in Nature*. Cambridge, Cambridge University Press, 1996.

Goldfield, E. C. *Emergent Forms: Origins and Early Development of Human Action and Perception*. New York, Oxford University Press, 1995.

Grusec, J. E. and Kuczynski, L. (eds.) *Parenting and Children's Internalization of Values: A Handbook of Contemporary Theory*. New York, John Wiley & Sons, 1997.

Hardin, R. *One for All: The Logic of Group Conflict*. Princeton, Princeton University Press, 1995.

Hare, R. M. *Essays in Ethical Theory*. Oxford, Clarendon Press, 1989.

Hare, R. M. *Essays on the Moral Concepts*. Berkeley & Los Angeles, CA; University of California Press, 1972.

Hare, R. M. *Freedom and Reason*. Oxford, Clarendon Press, 1963.

Hare, R. M. *The Language of Morals*. Oxford, Oxford University Press, 1964.

Hare, R. M. *Moral Thinking: Its Levels, Method, and Point*. Oxford, Clarendon Press, 1981.

Hare, R. M. *Sorting Out Ethics*. Oxford, Clarendon Press, 1997.

Harrison, G.A. and Boyce, A. J. (eds.) *The Structure of Human Populations*. Oxford, Clarendon Press, 1972.

Haworth, L. *Autonomy: An Essay in Philosophical Psychology and Ethics*. New Haven, Yale University Press, 1986.

Hegel, G. W. F. *Phenomenology of Spirit*, trans. A. V. Miller, Oxford, Oxford University Press, 1977.

Heidegger, M. *Being and Time*, trans. Macquarrie, J. and Robinson, E. London, SCM Press, 1962.

Heidegger, M. *The Piety of Thinking: Essays*, trans. J. G. Hart and J. C. Maraldo, Bloomington, Indiana University Press, 1976.

Heidegger, M. *Poetry, language, thought*, trans. A. Hofstadter. New York, Harper & Row, 1971.

Heidegger, M. *Pathmarks*, ed. W. McNeill. New York, Cambridge University Press, 1998.

Hostetler, J.A. *Amish Society*. Baltimore, Johns Hopkins, 1980.

Howells, C. *Sartre: The Necessity of Freedom*. Cambridge, Cambridge University Press, 1988.

Hudson, W. D. (ed.) *The Is-Ought Question: A Collection of Papers on the Central Problem in Moral Philosophy*. London, Macmillan, 1969.

Humberstone, I. L. "Direction of Fit." *Mind* 101, 1992, 59-83

Hume, D. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. P. H. Nidditch and L. A. Selby-Bigge. Oxford, Clarendon Press, 1975.

Hume, D. *A Treatise of Human Nature*, ed. L. A. Selby-Bigge. Oxford, Clarendon Press, 1975.

Huxley, T. H. and Huxley, J. *Evolution and Ethics, 1893-1943*. London, Pilot Press, 1947.

James, William. *Selected Writings*, ed. G.H. Bird. London & Vermont, Everyman, 1995.

Kahane, H. *Contract Ethics: Evolutionary Biology and the Moral Sentiments*. London, Rowman & Littlefield, 1995.

Kant, I. *Immanuel Kant's Critique of Pure Reason*, trans. N. Kemp Smith. London, Macmillan 1961

Kant, I. *Critique of Practical Reason*, trans. L. W. Beck. New York, Toronto, Macmillan, 1993.

Kant, I. *The Critique of Judgment*, trans. J. H. Bernard. Amherst, New York, Prometheus Books, 2000.

Kant, I. *Lectures on Ethics – Duties towards Animals and Other Spirits*, trans. L. Infeld. New York, Harper & Row, 1963.

Kerr, B. and Godfrey-Smith, P. "Individualist and Multi-level Perspective's on Selection in Structured Populations." *Biology and Philosophy* 17, 2002, 477-517.

Kierkegaard, S. R. *Either/Or*, vols. 1 & 2, trans. E. H. Hong and H. V. Hong. Princeton, Princeton University Press, 1987.

Killen, M. and Hart, D. (eds.) *Morality in Everyday Life: Developmental Perspectives*. Cambridge, New York; Cambridge University Press, 1995.

Kohlberg, L. *The Philosophy of Moral Development: Moral Stages and the Idea of Justice*. San Francisco; London, Harper & Row, 1981.

Kornblith, H. (ed.) *Naturalizing Epistemology*. Cambridge, Mass; London, MIT Press, 1994.

Kress, W. J. and Barrett G. W. (eds.) *A New Century of Biology*. Washington, DC, Smithsonian Institution Press, 2001.

Laupa, M. (ed.) *Rights and Wrongs: How Children and Young Adults Evaluate the World*. San Francisco, CA; Jossey-Bass, 2000.

Lemos, N. M. *Intrinsic Value: Concept and Warrant*. New York, Cambridge University Press, 1994.

Lewis, D. K. *Convention: A Philosophical Study*. Oxford, Basil Blackwell, 1986.

Loewy, E. H. *Freedom and Community: The Ethics of Interdependence*. Albany, State University of New York Press, 1993.

MacIntyre, A. *After Virtue: A Study in Moral Theory*. Notre Dame, Indiana; University of Notre Dame Press, 1981.

Mackie, J. L. *Ethics: Inventing Right and Wrong*. Harmondsworth; New York, Penguin, 1977.

Mackie, J. L. *Hume's Moral Theory*. London; Boston, Routledge & K. Paul, 1980.

Maienschein, J. and Ruse, M. (eds.) *Biology and the Foundation of Ethics*. Cambridge, Cambridge University Press, 1999.

Mandrell, J. *Don Juan and the Point of Honor: Seduction, Patriarchal Society, and Literary Tradition*. University Park, Pennsylvania; Pennsylvania State University Press, 1992.

Marx, K., Engels, F. and Lenin, V. I. *Critique of the Gotha Programme*, revised trans., ed. C P. Dutt. New York, International Publishers, 1966.

Maynard Smith, J. "Commentary on Kerr and Godfrey-Smith." *Biology and Philosophy* 17, 2002, 523-527.

Maynard Smith, J. *Did Darwin Get It Right?: Essays on Games, Sex and Evolution*. London, Penguin, 1993.

McGrew, W. C. *Chimpanzee Material Culture: Implications for Human Evolution*. Cambridge, Cambridge University Press, 1992.

Megarry, T. *Society in Prehistory: The Origins of Human Culture*. London, Macmillan, 1995.

Mill, J. S. *Utilitarianism; On Liberty; Considerations on Representative Government; Remarks on Bentham's Philosophy*, ed. G. Williams. London & Vermont, Everyman, 1993.

Mithen, S. *The Prehistory of the Mind*: *A Search for the Origins of Art , Religion and Science*. London, Thames & Hudson, 1996.

Moore, G. E. *Principia Ethica*, rev., ed. T. Baldwin. Cambridge, Cambridge University Press, 1993.

Nagel, T. *The View from Nowhere*. New York, Oxford University Press, 1986.

Narveson, J. "Alan Gewirth's Foundationalism and the Well-Being State." *Journal of Value Inquiry*, no.31.

Nietzsche, F. W. *Basic Writings of Nietzsche*, ed. W. A. Kaufmann. New York, Modern Library, 1992.

Nitecki, M. H. and Nitecki D. V. (eds.) *Evolutionary Ethics*. Albany, State University Of New York Press, 1993.

Nozick, R. *Invariances: The Structure of the Objective World*. Cambridge, Mass., Belknap Press of Harvard University Press, 2001.

Nozick, R. *The Nature of Rationality*. Princeton, Princeton University Press, 1993.

Nucci, L. P., Saxe G. B. and Turiel, E. (eds.) *Culture, Thought and Development*. Mahwah, NJ & London; Lawrence Erlbaum Associates, 2000.

Nussbaum, M. C. and Sen, A. K. *The Quality of Life*. Oxford, Clarendon Press, 1993.

Odling-Smee, F. J., Laland, K. N. and Feldman, M. W. *Niche Construction: The Neglected Process in Evolution*. Princeton, Princeton University Press, 2003.

Paradis, J., G. C. Williams and Huxley, T. M. *Evolution & Ethics: T. H. Huxley's Evolution and Ethics with New Essays on its Victorian and Sociobiological Context*. Princeton; Guildford, Princeton University Press, 1989.

Paul, E. F., Miller, F. A. and Paul, J. (eds.) *The Good Life and the Human Good* (Cambridge, Cambridge University Press, 1992.

Paul, E. F., Miller, F. D. Jr., Paul, J. and Ahrens, J. (eds.) *The New Social Contract: Essays on Gauthier*. Oxford, Blackwell, 1988.

Perry, R. *Galápagos*. Oxford; New York, Pergamon Press, 1984.

Piaget, J. *The Essential Piaget*. H. E. Gruber and J. J. Vonáeche (eds.), New York, Basic Books, 1977.

Piaget, J. *The Moral Judgment of the Child*. Routledge & Kegan Paul, 1932.

Pinker, S. *How the Mind Works*. New York, Norton, 1997.

Plato. *The Republic*, 2nd ed., trans. H. D. P. Lee. Harmondsworth, Penguin, 1974.

Platts, M. *Ways of Meaning*: *An Introduction to the Philosophy of Language*. London, Routledge & Kegan Paul, 1979.

Plotkin, H. C. (ed.) *The Role of Behavior in Evolution*. Cambridge, Mass., MIT Press, 1988.

Popper, K. R. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London, Routledge & Kegan Paul, 1974.

Popper, K. R. *The Poverty of Historicism*. London, Routledge & Kegan Paul, 1961.

Power, M. *The Egalitarians – Human and Chimpanzee*: *An Anthropological View of Social Organization*. Cambridge, Cambridge University Press, 1991.

Putnam, H. *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge Mass. & London, England; Harvard University Press, 2002.

Python Productions Ltd. *Monty Python's Big Red Book*. London, Eyre Methuen, 1972.

Quine, W. V. *From a Logical Point of View: Nine Logico-Philosophical Essays*. Cambridge, Mass., Harvard University Press, 1961.

Quine, W. V. and Ullian, J. S. *The Web of Belief*. New York, Random House, 1970.

Rawls, J. *A Theory of Justice*. Cambridge, Mass., Belknap Press of Harvard University Press, 1971.

Rawls, J. *Political Liberalism*. New York; Chichester, Columbia University Press, 1996.

Rawls, J. *The Law of Peoples; with, The Idea of Public Reason Revisited*. Cambridge, Mass., Harvard University Press, 1999.

Rescher, N. *A System of Pragmatic Idealism, Volume II: The Validity of Values*: *A Normative Theory of Evaluative Rationality*. Princeton, Princeton University Press, 1993.

Reed, E. *Encountering the World: Toward an Ecological Psychology*. New York, Oxford University Press, 1996.

Regis, E. (ed.) *Gewirth's Ethical Rationalism: Critical Essays with a Reply by Alan Gewirth*. Chicago, University of Chicago Press, 1984.

Richards, Robert J. *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*. Chicago and London; The University of Chicago Press, 1987.

Ridley, M. *The Origins of Virtue*. London, Viking, 1996.

Rogers, L. J. *Minds of Their Own: Thinking and Awareness in Animals*. St Leonards, N.S.W., Allen & Unwin, 1997.

Rogoff, B. *The Cultural Nature of Human Development*. Oxford and New York; Oxford University Press, 2003.

Rousseau, J.-J. *A Discourse on Inequality*, trans. M. W. Cranston. Harmondsworth, Penguin, 1984.

Sartre, J.-P. *Being and Nothingness: A Phenomenological Essay on Ontology*, trans. H. Barnes. New York, Washington Square Press, 1966.

Sartre, J.-P. *Notebooks for an Ethics*, trans. D. Pellauer. Chicago, University of Chicago Press, 1992.

Scanlon, T. *What We Owe To Each Other*. Cambridge, Mass., Belknap Press of Harvard University Press, 1998.

Schurz, G. *The Is-Ought Problem*: *An Investigation in Philosophical Logic*. Dordrecht, Netherlands; Kluwer Academic Publishers, 1997.

Seanor, D. and Fotion, N. (eds.) *Hare and Critics: Essays on Moral Thinking*. Oxford, Clarendon Press, 1988.

Sen, A. K. *Inequality Reexamined*. Cambridge, Mass., Harvard University Press, 1992.

Sen, A. K. "Well-being, Agency and Freedom: The Dewey Lectures 1984". *Journal of Philosophy* vol. 82, 1985.

Singer, P. *The Expanding Circle: Ethics and Sociobiology*. Oxford, Clarendon Press, 1981.

Singer, P. "The Triviality of the Debate over "Is-Ought" and the Definition of "Moral"." *American Philosophical Quarterly*, 1973, vol.10, pp.51-56

Skinner, B. F. *The Behavior of Organisms: An Experimental Analysis*. New York, Appleton-Century-Crofts, 1938.

Skinner, B. F. *Beyond Freedom and Dignity*. Harmondsworth, Penguin, 1973.

Skinner, B. F. *Science and Human Behavior*. New York; London, Free Press, Collier-Macmillan, 1953.

Skyrms, B. *Evolution of The Social Contract*. Cambridge, Cambridge University Press, 1996.

Smith, A. *The Theory of Moral Sentiments*, ed. D. D. Raphael and A. L. Macfie. Oxford, Clarendon Press, 1976.

Smith, A. *An Inquiry into the Nature and Causes of the Wealth of Nations,* vol. 2, ed. R. H. Campbell and A. S. Skinner. Oxford, Clarendon Press, 1976.

Sober, E. and Wilson, D. S. "Perspectives and Parameterizations: Commentary on Benjamin Kerr and Peter Godfrey-Smith's "Individualist and Multi-level Perspectives on Selection in Structured Populations." *Biology and Philosophy* 17, 2002, 529-537.

Sober, E. and Wilson, D. S. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, Mass., Harvard University Press, 1998.

Sterelny, K. *The Evolution of Agency and Other Essays*. Cambridge, Cambridge University Press, 2001.

Sterelny, K. *Thought in a Hostile World: The Evolution of Human Cognition*. Malden MA; Blackwell, 2003.

Stevenson, C. L. *Ethics and Language*. New Haven, Yale University Press, 1944.

Thagard, P. *Coherence in Thought and Action*. Cambridge, Mass.; London, MIT Press, 2000.

Thompson, P. (ed.) *Issues in Evolutionary Ethics*. Albany, State University Of New York Press, 1995.

Tiger, L. and Fox, R. *The Imperial Animal*. New York, Henry Holt and Company, 1989.

Trivers, R. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology*, vol. 4, pp.35-57.

Trivers, R. *Social Evolution*. Menlo Park, CA; Benjamin/Cummings, 1985.

Truman, H. S. *Off The Record: The Private Papers of Harry S. Truman*, ed. R. H. Ferrell. New York, Harper & Row, 1980.

Turiel, E. *The Development of Social Knowledge: Morality and Convention*. Cambridge, Cambridge University Press, 1983.

Turnbull, Colin. *The Mountain People*. London, Picador, 1974.

Upham, S. (ed.) *The Evolution of Political Systems*: *Sociopolitics in Small-Scale Sedentary Societies*. Cambridge, Cambridge University Press, 1990.

Vallentyne, P. (ed.) *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*. Cambridge, Cambridge University Press, 1991.

Voltaire. *The History Of Candide: Or, All For The Best*. London, Simpkin Marshall, n.d.

Von Neumann, J. and Morgenstern, O. *Theory of Games and Economic Behavior*. Princeton, Princeton University Press, 1964.

Waller, B. N. *The Natural Selection of Autonomy*. Albany, State University Of New York Press, 1998.

Weber, B. H. and Depew, D. J. (eds.) *Evolution and Learning: The Baldwin Effect Reconsidered*. Cambridge, Mass., MIT Press, 2003.

Williams, B. A. O. *Ethics and the Limits of Philosophy*. London, Fontana, 1985.

Williams, G. C. *Adaptation and Natural Selection: A Critique of Some Evolutionary Thought*. Princeton, NJ; Princeton University Press, 1966.

Williams, G. C. *Natural Selection: Domains, Levels, and Challenges*. New York, Oxford University Press, 1992.

Wilson, E. O. *In Search of Nature*. Washington, D.C., Island Press, 1996.

Wilson, E. O. *On Human Nature*. Cambridge, Mass.; London, Harvard University Press, 1978.

Wilson, R. A. (ed.) *Character Above All: Ten Presidents from FDR to George Bush*. New York; London, Simon & Schuster, 1995.

Wittgenstein, L. *Philosophical Investigations*, trans. G. E. M. Anscombe. Oxford, Basil Blackwell, 1968.

Wozniak, R. H. and Fischer, K. W. (eds.) *Development in Context: Acting and Thinking in Specific Environments*. Hillsdale, N.J., L. Erlbaum, 1993.

Wright, R. *Nonzero: The Logic of Human Destiny*. London, Little Brown, 2000.

Zimbardo, P. G. *Psychology and Life*, 13th ed. New York, HarperCollins, 1992.

Zimmerman, M. J. *The Nature of Intrinsic Value*. Lanham and Oxford; Rowman & Littlefield, 2001.

NOTES

INTRODUCTION

John Rawls, *A Theory of Justice* (Cambridge, Mass., Harvard University Press, 1971)

[2] David Gauthier, *Morals by Agreement* (Oxford, Clarendon Press, 1986)

[3] Daniel Dennett, *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (New York, Simon & Schuster, 1995)

[4] Bruce Waller, *The Natural Selection of Autonomy* (Albany, N.Y., SUNY Press, 1998)

[5] e.g. Baldwin, Waddington, Mayr, Lorenz, Piaget – see *The Role of Behavior in Evolution*, ed. H. C. Plotkin, Cambridge, MIT Press, 1998); See also *Evolution and Learning: The Baldwin Effect Reconsidered*, ed. B. H. Weber & D. J. Depew (Cambridge, Mass., MIT Press. 2003)

[6] Alan Gewirth, *Reason and Morality* (Chicago, University Of Chicago Press, 1978)

[7] see section (ii) of Chapter Five

[8] see section (ii) of Chapter Six,

[9] Cf. Peter Singer, *The Expanding Circle: Ethics and Sociobiology* (Oxford, Clarendon Press, 1981)

[10] Cf. M. C. Nussbaum & A. K. Sen eds. *The Quality of Life* (Oxford, Clarendon Press, 1993)

CHAPTER ONE