

Psychology and Ethical Epistemology: An Ash‘arī Debate with Mu‘tazilī Ethical Realism, 11th-12th C.

AYMAN SHIHADDEH (SOAS University of London & New York University Abu Dhabi)

Abstract

We examine a hitherto unstudied debate, turning on the epistemology of value judgements, between Ash‘arīs and Baṣran Mu‘tazilīs of the late eleventh and twelfth centuries. Al-Ghazālī and al-Rāzī countered Mu‘tazilī ethical realism, here defended by al-Malāḥimī, by developing an emotive subjectivism underpinned by increasingly sophisticated psychological accounts of ethical motivation. Value judgements, they maintained, arise not from knowledge of some ethical attributes of acts themselves, but from subjective inclinations, which are often elusive because they can be unconscious or indirect. We also argue against the widespread notion that Ash‘arīs espoused an anti-rationalist ethics, and we show that they were not only ethical rationalists, but also the more innovative side in this debate.

Keywords: al-Ghazālī, Fakhr al-Dīn al-Rāzī, al-Malāḥimī, Avicenna, Ash‘arism, Mu‘tazilism, Value theory, Moral realism and anti-realism, Emotivism, Moral psychology, Rationalism, Intellect (*‘aql*), Estimation (*wahm*), Disposition (*tab*), Widely-accepted premises (*mashhūra*), Reputable premises (*maḥmūda*)

Introduction

Two main metaethical theories were advanced in medieval Islamic theology and jurisprudence. The Baṣran Mu‘tazila upheld an ethical realism, according to which ethical value is a real and intrinsic attribute of the evaluable act, and thus, like other attributes inhabiting the external world, is necessarily knowable.¹ Value judgements passed on acts, as in ‘Charity is good’ and ‘Murder is bad’, accordingly articulate knowledge of the external world. Classical Ash‘arīs responded with an anti-realist metaethics, on the back of which they advocated a theological voluntarism. This anti-realism was coupled with an increasingly sophisticated subjectivism, according to which ethical value is ordinarily tied to an act’s subjective consequences for an agent, measured ultimately in terms of the agent’s emotive states. This position eventually culminated in the rise of a consequentialist normative ethics. A considerable amount of attention has been afforded to these theories in recent scholarship. Most relevant to the present article is a 1976 study by George F. Hourani on al-Ghazālī’s (d.

¹ The latter view follows from their general epistemological conviction that if something is real it must be knowable to us (SHIHADDEH 2013).

505/1111) ethics of action and our own 2006 monograph on Fakhr al-Dīn al-Rāzī's (d. 606/1210) ethical theory and more recent work on classical-Ash'arī criticism of Mu'tazilī ethics (HOURANI 1976; SHIHADDEH 2006; 2016: 396-40).²

One of the broad aims of the present article is to challenge the widespread notion that Ash'arīs espouse an 'anti-rationalist' ethics.³ This notion is in the first instance prompted by their own self-description as opponents to the ethical theory of the Mu'tazila, whose conception of ethical value they characterise as rational (*'aqlī*). It may seem fairly uncontroversial hence to cast them, apparently by their own admission, as anti-rationalists. The implication of invoking this rather nebulous dichotomy of rationalism and anti-rationalism in this context is that the Mu'tazila undertook comparatively sophisticated and innovative theorisation in ethics, whereas Ash'arīs were simply opposed to any sort of theorisation in this area on the grounds that it lay beyond human understanding. Against this characterisation, the present article will argue that the ethical thought of Ash'arīs is in no way anti-rationalist, and that in the last phase of the debate between them and the Mu'tazila they in fact were theoretically and dialectically the more sophisticated, innovative and resourceful side. This is to say that as well as gaining the political support that cemented their ascendancy over their adversaries, Ash'arīs also won the dialectical battle. Light will be shed on this phase mainly through a hitherto unstudied debate between the two sides, involving the two aforementioned Ash'arīs, al-Ghazālī and al-Rāzī, as well as the intermediating Baṣran Mu'tazilī Rukn al-Dīn al-Malāḥimī (d. 536/1141), a younger contemporary of al-Ghazālī. Al-Malāḥimī's contribution to this debate occurs in his partially-extant theological work *al-Mu'tamad fī uṣūl al-dīn*, the discussion on ethics in which became available in print only after I published my book on al-Rāzī's ethics.⁴ He responds to al-Ghazālī's refutation of Mu'tazilī ethical theory, and is in turn responded to by al-Rāzī.

The central problem in the debate is whether value judgements express knowledge of mind-independent reality or are grounded in subjective inclination. The latter view was associated in earlier *kalām* with what we may describe as a simple self-centred subjectivism, according to which an act is good if it is pleasurable to the agent, bad if it results in pain. While it was fairly easy for the ethical realists of classical *kalām* to counter this position, we shall see that, partly under philosophical influence, al-Ghazālī and al-Rāzī developed more complex forms of subjectivism through innovative psychological accounts of motivation, arguing that the self-centred motives that evoke value judgements could be unconscious or indirect and could give rise to widely-accepted ethical rules, which are *prima facie* non-self-centred. The two Ash'arīs approach the problem differently, with al-Ghazālī exhibiting the influence of Avicenna's (d. 428/1037) treatment of widely-accepted premises (*mashhūrāt*) and his faculty psychology, particularly the faculty of estimation—an aspect left out in

2 On Mu'tazilī ethics, see SHIHADDEH 2016: 391-6 (on their theory of ethical value); VASALOU 2008; HOURANI 1971. See now also VASALOU 2016, on the ethical theory of Ibn Taymiyya and its earlier background.

3 For instance, MADELUNG (2015: 28-9; cf. 26), where Ash'arī theology as a whole is said to be "predominantly anti-rationalist". But see now BOUHAFI 2021 and SYED 2016, who challenge this notion, albeit from an angle different from the one taken here.

4 The *Mu'tamad* has been published in two editions, both incomplete: the first, edited by Wilferd MADELUNG and Martin MCDERMOTT, appeared in 1991, and the second, edited by Madelung, appeared in 2007. Only the latter edition contains the discussion on ethical value.

Hourani's article—and al-Rāzī shifting away from faculty psychology and into the domain of social psychology. We shall also see how al-Malāḥimī, the pre-eminent Mu'tazilī at the time, failed to keep up with his Ash'arī adversaries.

Our starting point, however, will be the aforementioned pre-Ghazālian, classical-Ash'arī opposition to the purportedly 'rationalist' ethics of the Mu'tazila. We shall argue that the term 'aqlī refers here not to rationalism, but to realism, and that what classical Ash'arīs advocated, therefore, was not anti-rationalism, but ethical anti-realism and subjectivism. As a consequence of a subsequent shift in the theological conception of 'aql, al-Rāzī declares the subjectivist account of ethical value to be just as rational as the realist one.

1. Classical-Ash'arī Ethical Anti-Realism and Conceptions of Intellect ('Aql) and Disposition (Tab)

The debate between the two sides centres, in the first place, on the nature and grounds of ethical value.⁵ The Baṣran Mu'tazila champion a realist position, whose central tenet is that ethical values are real attributes (*ṣifa*) of acts and hence inhabit the external world. They are not caused by the essences of acts, but rather by act-configurations (*wajh*), which are specific combinations of circumstances that may accompany certain types of acts (SHIHADAH 2016: 391-6). These circumstances can be specific aspects of the agent, the patient, or the wider state of affairs. For instance, if (1) a series of speech acts have the form of a statement, and (2) the speaker has the intention to convey a statement (and so, for example, is not speaking in his sleep, or under compulsion), and (3) the statement does not correspond to a true state of affairs and is hence false, then the speech acts will have the configuration that we normally label as 'a lie'. This configuration renders the speech act morally bad. As al-Malāḥimī writes:

[Bad acts] are bad on account of configurations that characterise them when they occur (*wujūh taqa'u 'alayhā*), which is to say that when [an act] occurs it coincides with specific circumstances (*qarīna*) that can be either negations or affirmations. For instance, when the occurrence of harm coincides with [*a*] it being undeserved, [*b*] it preventing no [greater] harm, be it certain or probable, and [*c*] it [serving] no sound purpose, it will be bad. [Our school members] express this by saying, "[This given harmful act] is bad because it is wrongful (*ẓulm*)". Another example is when the occurrence of a belief coincides with [the fact that] the object of the belief is contrary to how it is believed to be. They refer to this as 'misbelief' (*jahl*), and assert, "[This

⁵ The central value judgements are 'good' (*ḥasan*), defined as that for the performance of which the agent deserves no blame, and 'bad' (*qabīh*), defined as that for the performance of which the agent deserves blame. Three further judgements are subdivisions of the judgement 'good': 'permissible' (*mubāh*, that for neither the performance nor non-performance of which the agent deserves either praise or blame), 'recommended' (*mandūb*, that for the performance of which the agent deserves praise, but for the non-performance of which deserves no blame), and 'obligatory' (*wājib*, that for the performance of which the agent deserves praise, and for the non-performance of which deserves blame). These, al-Malāḥimī explains, are the "principal judgements that apply to acts (*al-uṣūl fī aḥkām al-af'āl*); and all other [ethical] attributes of acts ultimately reduce to them" (2007: 831). For instance, 'prohibited' (*maḥzūr*, *muḥarram*) reduces to 'bad'.

given belief] is bad because it is a misbelief". They do the same with all other bad acts. (al-MALĀHIMĪ 2007: 851; cf. 2016: 168)

Connected to this ontological contention is an epistemological one, namely, that the ethical attributes of acts are knowable to the mind without the aid of revelation or tradition. The ethical attributes of certain types of act are self-evident, and hence known immediately (*darūrī*) to all sound-minded human beings, while the ethical attributes of other types of act are not self-evident but can be acquired through inference (*nazarī*).

Against this theory, classical Ash'arīs, as I showed elsewhere, counter with an anti-realist position, arguing that when value expressions are encountered in ordinary language—that is, not in the specialised, religious-conventional sense of being commanded or prohibited by God—they are grounded not in the extra-mental reality of acts themselves, but in the subjective experience of attraction and repulsion, which arise from the disposition (*tab'*) of an individual in reaction to things.⁶ According to al-Juwaynī (d. 478/1085) and other school-members, the lexical definition of 'good' is pleasurable and beneficial, that of 'bad' painful and harmful (al-JUWAYNĪ 2010: II, 732). What classical Ash'arīs advance, therefore, is an emotivist subjectivism, which they invoke by and large dialectically to refute the central metaethical thesis of Mu'tazilī realism, and hence to present an account of value that is confined to the metaethical plane and does not motivate a normative, or prescriptive ethics. Their metaethical position was instead supplemented by the normative view that divine command is the only non-subjective, and hence authoritative, source of value judgements. An upshot of this view is that God's own acts are not subject to ethical rules.

One aspect of this account that has so far eluded serious analysis is that classical Ash'arīs situate their anti-realist ethics in direct opposition to the 'rational' conception of ethical value (*al-ḥusn wa-l-qubḥ al-'aqliyyān* or *taḥsīn al-'aql wa-taqbīḥu-hu*) of the Mu'tazila. How should we understand this ostensible opposition to rationalist ethics, especially when classical Ash'arīs themselves make no appeal to revelation in their subjectivist ethics, which hence appears just as deserving of the label 'rational'? The first thing to note is that in this classical-Ash'arī formulation of the Mu'tazilī position, what the adjective 'rational' qualifies is not their adversaries' ethical theory (their *qawl* or *madhhab*), but ethical value: it is goodness and badness that are rational, according to the Mu'tazila, not the Mu'tazilī theory of goodness and badness. So in what sense is ethical value rational or non-rational? The background to this question, which to my knowledge has hitherto remained unexplored, should be sought in the classical-Ash'arī conception of 'intellect', particularly in the mature position of al-Bāqillānī (d. 403/1013).⁷ To him, the expression 'intellect' (*'aql*) refers, not to a cognitive faculty (*quwwa*), but to a body of immediate knowledge-items (*'ulūm darūriyya*) that differentiate rational beings from non-rational living beings, including humans of unsound intellect, such as the insane and children, and non-human animals. To be of sound intellect (*'āqil*), or *compos mentis*—an important notion in Islamic law—one must be possessed of these knowledge-items in full. Al-Bāqillānī identifies the knowledge that constitutes the intellect firstly by eliminating two subdivisions of immediate knowledge that

6 SHIHADĒH 2016: 399–401. Acts in classical *kalām* are, properly speaking, things, specifically accidents (as understood in *kalām* atomism, of course).

7 A similar position is attributed to al-Ash'arī. A full investigation of this conception goes well beyond our present purview and will be undertaken in a future study.

do not meet the criterion just identified. Sensory knowledge and introspective knowledge (literally, knowledge of objects occurring within oneself, *fi l-nafs*)—both types of subjective immediate knowledge—are shared with humans of unsound intellect and animals, and hence fall outside the scope of the intellect (al-BĀQILLĀNĪ 1998: I, 197; 188-90). Introspective knowledge includes, for example, knowledge of the pleasure, pain, desire, repulsion, motivation and will that occur within the knower, and even, according to al-Bāqillānī, one's knowledge of the existence of oneself. Three further subdivisions of immediate knowledge, on the other hand, are not shared with humans of unsound intellect and animals, and hence together define the scope of the intellect (al-BĀQILLĀNĪ 1998: I, 196-7; 190-2; cf. 1957: 10-11). The first, and most relevant here, is self-evident (*badīhī*) knowledge, which includes what we may describe as logical truths—for instance, a thing cannot simultaneously both exist and not exist; two is more than one; and two contrary things cannot be co-located.⁸ Al-Bāqillānī delineates self-evident knowledge only through examples, but his successor al-Juwaynī limns it in more general terms as comprising the knowing of certain impossibilities as impossible, and of certain possibles as possible (al-JUWAYNĪ 1996: I, 112-13). Crucially, al-Juwaynī does not include knowledge of all impossibilities and possibilities, because much of this knowledge is inferential.

With this definition of intellect, immediate knowledge-items are described as rational, or as deriving from the intellect (*fi l-'aql*), only if they fall within the scope of the intellect just described. Their objects are all extra-mental, necessary in themselves, and hence inalterable.⁹ So they will be recognised by all people of sound intellect as true. By contrast, sensory and introspective knowledge is available only to the individual knower, be it a person or an animal, and is as such subjective. What is more, it is unnecessary and alterable; for instance, pain does not occur necessarily following injury, but is dependent on God's will. So despite being knowledge (*'ilm*) in the fullest sense, and certainly no less so than self-evident knowledge, sensory and introspective knowledge is not 'rational' (*'aqlī*), in the sense of deriving from the intellect. Classical Ash'arīs characterise some of these internal objects of introspective knowledge as being grounded in an individual's disposition (*tab'*), which refers to the extent to which one has desire (*shahwa*) for certain perceivable things or types of things, and is hence predisposed to find pleasure in perceiving them, or repulsion (*nafra*, *nufūr*, *nifār*) from other things or types of things, and is hence predisposed to find pain in perceiving them.¹⁰ So the disposition, as classical Ash'arīs insist against philosophical accounts thereof, is not a thing in itself, but, as al-Ash'arī puts it, the ordinary occurrence of "certain accidents within certain bodies" (IBN FŪRAK 1987: 132; 279). By this occasionalist account of the temperament, if one person (body A) has revulsion towards certain things he

8 The two other types of immediate knowledge that al-Bāqillānī considers to comprise the intellect are knowledge of the normal course of events (*'āda*), which God preserves (such as knowing that certain things will inevitably, though not necessarily, burn when they come into contact with fire and the meanings intended by a speaker), and knowledge of the objects of widely-transmitted reports (*tawātur*), including past occurrences and remote places. The former type should be read against the backdrop of classical-Ash'arī occasionalism. These types of knowledge are of little relevance to our present purposes.

9 In this vein, I previously translated the *kalām* term *'illa 'aqliyya* (which is contrasted with *'illa sam'iyya*) as 'real cause', rather than 'rational cause' (SHIHADÉH 2013: 204). The latter is a literal but erroneous rendering; for despite being seemingly epistemological, the term is firmly grounded in ontology, specifically in the notion that this type of cause lies in the reality (*ḥaqīqa*, pl. *ḥaqā'iq*) of things.

10 See the Mu'tazilī 'ABD AL-JABBĀR 1965: 17 ff.

will be predisposed to become angry whenever he perceives them (revulsion and perception being explained as accidents supervening upon his body); and if another person (body B) has desire for certain things he will be predisposed to experience pleasure whenever he perceives them, and so forth. Some of these predispositions may be common to all or most humans, albeit to different degrees; others are specific to individuals. Some people, al-Ash'arī reportedly observes, have a benevolent disposition, while others have a malevolent disposition (IBN FŪRAK 1987: 132).

So when classical Ash'arīs reject the Mu'tazilī characterisation of goodness and badness as rational, they only deny that they are extra-mentally real attributes of acts, that the ethical attributes of some acts are self-evident, and consequently that those of others are inferred from those known immediately. This stance does not equate to an outright denial of the 'rationality' of these concepts, in a fashion that would render them bereft of mentally cognisable referents in the absence of revelation. All immediate knowledge-items are cognisable in this way. When classical Ash'arīs then say that goodness and badness arise from the disposition, they mean to ground ethical value in internal perceptions, particularly in what we would call emotions. When an individual observes certain acts, this perception may be followed by a sensation of pleasure or pain, or attraction or repulsion, which would be knowable to him, and he may accordingly describe those acts as good or bad. This is an emotive account of ethical value, not an anti-rational one.

2. Al-Ghazālī: Value Judgement and the Errors of the Estimation

The classical-Ash'arī emotive, anti-realist position is developed further by al-Ghazālī under the philosophical influence of Avicenna. His most extensive discussion occurs in his juristic work, the *Mustasfā*; a slightly shorter version of the discussion is offered in the theological work, the *Iqtisād* (al-GHAZĀLĪ n.d.: I, 177-94; 1962: 160-74). Al-Ghazālī explains the widely-accepted, conventional senses (*al-iṣṭilāḥ al-mashhūr al-āmmī*) of 'good' and 'bad' in ordinary language, respectively, as "agreement (*wāfaqa*) with the agent's ends (*gharaḍ*)" and "disagreement (*khālafā*) with the agent's ends"—in other words, convenience and inconvenience to the agent.¹¹ The same act, hence, can be good in relation to one person, and bad in relation to another. These evaluations stem from the attraction (*mayl*) and repulsion (*nafra*) that arise from the agent's disposition in reaction to things and acts, and they are no different than the attraction or repulsion that one may experience, say, when seeing attractive or unattractive visual human forms.

Having introduced his subjectivist definitions of the central value terms, al-Ghazālī refutes three claims that he attributes to the Mu'tazila, the first ontological, the second and third epistemological: (1) that goodness and badness are essential attributes (*wasf dhātī*) of acts; (2) that certain value judgements are self-evident knowledge-items; and (3) that all sound-minded people agree on self-evident value judgements, which confirms their self-evidence (al-GHAZĀLĪ n.d.: I, 182-3). By focusing on value judgements purported to be self-evident to the exclusion of those inferred from them, al-Ghazālī follows in the footsteps of

¹¹ AL-GHAZĀLĪ n.d.: I, 179-81; cf. HOURANI 1976; MARMURA 1969. Other studies have dealt with the subject but not added much.

his teacher al-Juwaynī, who argues that once the falsity of the former is exposed, that of the latter will automatically follow (al-JUWAYNĪ 1950: 259-60). Al-Ghazālī refutes the first, ontological claim with ease: the same act, he reasons, can be good in some cases, bad in others—a case in point being inflicting harm on another human—and therefore can be essentially neither good nor bad, because essential attributes are inalterable (al-GHAZĀLĪ n.d.: I, 183). As we shall see in the next section, this is a strawman argument. The second, epistemological claim is rejected on the grounds that the value judgements that the Mu‘tazila claim to be self-evident are in fact not accepted by many non-Mu‘tazilīs (al-GHAZĀLĪ n.d.: I, 183-4).

Al-Ghazālī’s response to the third claim is the lengthiest and most important part of the discussion. The problem he tackles is this. If value judgements, as he and other Ash‘arīs claim, are not grounded in reality, then why are they considered by all (or at least the vast majority of) sound-minded people to be self-evident truths and epistemically on a par with (genuine) self-evident truths? How could a false belief—for instance, that lying is intrinsically bad—be apprehended as self-evident knowledge, not by a minority of misguided individuals, but by all (or the overwhelming majority of) people? Al-Ghazālī does not hesitate to concede that people do in fact come to agree unanimously, or almost unanimously, on untrue beliefs and even to construe them as self-evident knowledge. And like earlier Ash‘arīs, he insists that widely-accepted value judgements are little more than social conventions (‘urf) (SHIHADĒH 2016: 399-400). However, al-Ghazālī goes further than earlier Ash‘arīs by offering an explanation of how these judgements become widely-accepted conventions, and this he does by applying aspects of Avicenna’s treatment of widely-accepted premises (*mashhūrāt*) and his theory of the psychological faculty of estimation (*wahm*).

Al-Ghazālī’s deep interest in Avicenna’s account of the causes of widely-accepted premises is well-attested in another, logical work, *Mi‘yār al-‘ilm*, where this account is both adapted and developed (al-GHAZĀLĪ 1961: 193-7; MARMURA 1969: 393-6). For Avicenna, ethical premises, which he sometimes terms reputable (*maḥmūda*) premises, are acquired, as opposed to innate, widely-accepted premises.¹² They derive from social conventions and become deeply embedded (*mutaqarrira*) in individuals, so much so that ordinary people may deem them epistemically equivalent to primary premises (IBN SĪNĀ 1938: 63). Avicenna does not explain, in his discussion of reputable premises, how this embeddedness occurs, but he indicates elsewhere that it happens through the psychological faculty of estimation becoming conditioned to ethical conventions such that it issues emotive judgements on their objects (BLACK 1993: 243-4). Several causes that give rise to these conventions are mentioned briefly in the *Najāt*, including the desire for peace-making and conciliation, and ancient laws (*sunan qadīma*) that survived from obsolete systems of belief and practice (IBN SĪNĀ 1938: 63).

12 IBN SĪNĀ 1882: 58-9; 1938: 63-4; 1956: 65-6. Innate (*fīṭrī*) widely-accepted premises include primary (*awwalī*) principles of reason and estimative premises, which should be identified more precisely as ‘pure estimative’ premises (*wahmiyyāt ṣirfa*). On reputable premises in general and the background to Avicenna’s treatment thereof, see BLACK 1990: 95-101; AOUAD 1997. Rendering *qadīyya maḥmūda* as ‘reputable premise’ captures the senses of being widely accepted and held in high esteem, which is how this type of premise is characterised in Arabic logical sources. The term derives from the Greek *endoxon*, which can be rendered as ‘reputable opinion’ (on the Greek background of the term, see CELLI 2018: 98-100). The standard rendering of the Arabic term as ‘praiseworthy premise’ conveys a prescriptive sense, absent in the Arabic expression, but not the sense of wide acceptance, and is therefore inadequate.

Al-Ghazālī's account of the causes of value judgements in the *Mustasfā* draws heavily on Avicenna's account and develops it in various ways, most importantly by laying emphasis on the role of the estimation. Value judgements, he says, are reputable, widely-accepted premises (*qaḍāyā maḥmūda mashhūra*), which become widely-accepted in either of the following two ways (al-GHAZĀLĪ n.d.: I, 186).

The first is that some judgements originate in the teachings of revelation (al-GHAZĀLĪ n.d.: I, 184-5). The ethical rules stipulated by revealed religion are accepted by those among their adherents who have direct access to those teachings, and they are then disseminated more widely among those who follow them uncritically (*taqlīd*). As al-Ghazālī's logical works make clear, the former group receive those teachings in the form of premises accepted on the basis of either wide transmission (*tawātur*) or authority (*maqbulāt*), and therefore not as reputable premises (al-GHAZĀLĪ 1925: 52; 1961: 197-8). So they are received as reputable premises only by the class of ordinary people, in whom the ethical teachings of religion are inculcated from childhood to the extent that they become dissociated from their religious roots and viewed as self-evident truths (al-GHAZĀLĪ 1961: 196). Although Avicenna does not count dominant living religions (such as Islam in Muslim-majority societies) among the sources of widely-accepted premises, al-Ghazālī clearly takes his cue from several elements of the philosopher's treatment of these premises, most probably including the reference to ancient belief systems.¹³

The second way is that many value judgements arise and become widely established among people on account of the extent to which they further or hinder the ends (*gharaḍ*) of individuals. Which is to say that they are grounded in the subjective consequences of acts. Ends here are understood to be the objects sought by an agent's will (*irāda*) for the purpose of fulfilling a need of the agent. According to the Mu'tazila, these are not only conscious, which is to say that they are known to the agent (unlike desires [*shahwa*], whose presence is often unknown to the agent [IBN MATTAWAYH 2009: II, 414]), but also voluntary. And, of course, they refute the notion that value terms originate in the will of any agent, be it directed at self-centred ends or otherwise ('ABD al-JABBĀR 1962: 81 ff.). As al-Ghazālī notes, the Mu'tazila would counter his explanation by adducing value judgements that, they contend, do not arise from self-centred subjectivist considerations, such as the obligations to tell the truth, to assist those in severe need and to keep secrets and promises (al-GHAZĀLĪ n.d.: I, 185-6).

Yet he insists that even widely-accepted value judgements that appear not to be self-centred stem from subjective ends, except that the subjective ends that underpin them are very elusive (*tadiqqu wa-takhfā*) and can only be discerned by critical investigators (*muhaqqiq*)—that is, those who are both highly learned and skilful in independent, critical thinking to the extent that they are capable to navigate difficult problems and arrive at the

¹³ Al-Ghazālī's view that revealed religions are a source for non-religious ethical maxims current among ordinary people is complemented by his claim that the elite virtue ethics of the philosophers (*al-falsafa al-khuluqiyya*) has its origins in Sufism—presumably not only Islamic Sufism but also ethico-mystical traditions in pre-Islamic revealed religions (1969: 24). Revelation is thus a major source for the non-religious ethics of people of all educational strata.

truth (*haqq*).¹⁴ Most scholars fall below this rank (al-GHAZĀLĪ n.d.: I, 190; cf. 1925: 57). This is to say that these value judgements are reputable premises, whose causes—namely, the subjective ends of people—are not ordinarily known to those who uphold them, but can be identified through in-depth investigation. Al-Ghazālī unpacks his claim along psychological lines, setting out what he describes as three “sources of error (*ghalat*)”, through which value judgements obtain in people’s minds and become misconstrued as self-evident truths about the external world.¹⁵ All three stem from the pseudo-rational estimative faculty (*wahm*), a component of Avicenna’s theory of the psychological faculties, which al-Ghazālī adopted.¹⁶ He writes:

The psychological faculties of most people obey these false estimations, even when they know their falsity. The acting and abstention from acting of most people are due to these estimations; for estimation has a great hold on the soul.¹⁷

The role of the estimative faculty, which al-Ghazālī does not spell out here, is that it issues emotive judgements on objects in the external world, presenting them as though they were self-evident objective truths apprehensible to reason, and that it is prone to error in its judgements.

The first type of error is that evaluations that are subjective and relative are conceived as objective and absolute. Individuals use the expression ‘bad’ for what is contrary to one’s own personal ends even if it agrees with the ends of others, and ‘good’ for what agrees with one’s personal ends even if it is contrary to the ends of others, but then project these evaluations externally onto objects, thus perceiving them as non-sensible properties intrinsic to the objects themselves. The relative goodness or badness of acts thereby becomes perceived as absolute goodness or badness. When individuals commit this error, they pay no heed to how an act affects others, or even neglect to consider how the same type of act affected them personally on previous occasions in the past or may affect them in the future.

The second is that subjective considerations tend to give rise to simple universal rules through a process of incomplete induction (cf. al-GHAZĀLĪ 1961: 196). If something is disadvantageous most of the time, it will be judged by the estimation to be absolutely bad, even if and when it is advantageous or imperative in a minority of cases. Individuals become habituated to those rules and do so to the extent that they find in themselves the urge to adhere to them and loathing towards breaking them. The rule becomes a subjective end in itself, such that adhering to it is deemed to be in agreement with the agent’s ends and hence good, and failing to do so is deemed to be in disagreement with his ends and hence bad. Al-Ghazālī gives the example of the absolute badness of lying, which is inculcated in children to the extent that it becomes deeply ingrained in them. To avoid undermining their abhorrence of lying, children are not informed that in some instances lying is in fact good, such as lying to

14 Al-GHAZĀLĪ n.d.: I, 185-6; cf. I, 190, where those scholars are described as having been afforded knowledge of truth by God. It would obviously make little sense to translate *muhaqqiq* here as ‘verifier’, as seems to be the trend these days.

15 Al-GHAZĀLĪ n.d.: I, 187-90; 1962: 166-9. Al-Ghazālī’s usage of the term ‘error’ in logical and epistemological contexts is a philosophical influence.

16 On estimation in Avicenna and al-Ghazālī, see, respectively: BLACK 1993, and GRIFFEL 2012.

17 Al-GHAZĀLĪ n.d.: I, 190, apparently echoing IBN SĪNĀ 1959: 167, 182-3.

save the life of a virtuous person. This, however, explains how lying comes to be viewed as bad by those in whom this view is inculcated, but leaves unexplained, in the framework of al-Ghazālī's subjectivism, why such a view would be inculcated in the first place. (Al-RĀZĪ, as we shall see, addresses this question directly.) It is arguable that when parents and others impress upon children the badness of lying, they do so to guard against the subjective detriment that individuals expect from interacting with dishonest people, especially close family members.

The third type of error is that if something restricted (*khāṣṣ*), X, is always associated (*maqrūn*) with something more general (*ʿāmm*), Y ('Every X is a Y'), the estimation will be inclined, incorrectly, to convert (*'aks*) this relation, thus yielding the conviction that Y is invariably associated with X ('Every Y is an X'). Through this false conversion, if X is repulsive, Y becomes repulsive by association: as al-Ghazālī writes, "all that is associated with something pleasurable itself becomes pleasurable, and all that is associated with something detestable itself becomes detestable".¹⁸ Several examples are given. For instance, if someone is bitten by a snake, he will be repulsed by any object whose shape and colour resemble those of a snake, such as a patch-covered rope; so from 'Every snake is a patch-covered rope-like object' and 'Every snake is repulsive', one deduces, 'Every patch-covered rope-like object is repulsive'. And if a common person is opposed to a school of thought on account of certain doctrines of theirs, he will reject any doctrine attributed to them, even though he may accept it if it is presented to him without it being attributed to that school.

Having set out these errors, al-Ghazālī goes on to explain how seemingly non-self-centred value judgements arise and how people become motivated to abide by them. He starts with the purported obligation to save a person on the verge of death, if one is easily able to, which the Mu'tazila adduce as an example of a self-evident obligation (al-GHAZĀLĪ n.d.: I, 190-1). Al-Ghazālī responds by explaining this judgement in ways that map onto the three types of error he has just set out. The imperative to provide assistance, he contends, originates above all in the pain caused by the inborn disposition known as "tenderness associated to the genus" (*riqqat al-jinsiyya*), which is triggered at the agent's perception of the suffering of another human, or sometimes a beast.¹⁹ This, in other words, is the disposition of sympathy. When an agent encounters a human on the verge of death, he imagines himself in that person's place and imagines other people refusing to assist and he finds their refusal reprehensible; he then imagines the dying person having these same thoughts about him, and becomes distressed by that. To alleviate the distress caused by these self-centred imaginations, the agent becomes motivated to assist that person. Al-Ghazālī here appears to apply the first type of error: the agent does not recognise the self-centred end motivating his act, but in most cases will think that the goodness of the act is intrinsic and absolute. But what if the agent feels no sympathy

18 Al-GHAZĀLĪ n.d.: I, 191. This echoes Avicenna's remark that one may be disgusted by honey because of its similarity to bile, which he attributes to the activity of the estimation (1959: 182).

19 Al-GHAZĀLĪ n.d.: I, 190; cf. 1961: 193-5. The term often appears as *al-riqqa al-jinsiyya*. The expression *jins* here is originally intended in the *kalām* sense of 'class', a reference to either humans or animals—whence the expression *abnā' jinsihi* ('members of his/its kind'). Al-Ghazālī and later theologians may have intended the philosophical sense of 'genus', which would be a reference to animals. In practice, however, not all animals are objects of sympathy to the same degree. Humans tend to sympathise more with horses than with mice, and much less or not at all with insects.

for the dying beast or person he encounters? Al-Ghazālī says that it is hard to imagine an individual experiencing no sympathy whatsoever, but that even if this were granted another motive would remain operative—namely, the expectation of praise from others for a praiseworthy act. This would be an instance of the second type of error, whereby a widely-accepted universal rule becomes ingrained in individuals through social habituation. If the agent offers assistance but does not expect others to observe or know his praiseworthy act, the act would then be motivated by the mental association between the act of assisting those in dire need of help and praise. One is habituated to associate this sort of act with praise, and thus comes to think that the act is praiseworthy in absolutely all instances, when in fact it receives praise only in the presence of others who may praise it. Al-Ghazālī here explicitly appeals to the third way in which estimation engenders value judgements.

Al-Ghazālī, therefore, advances a subjectivist account of value, which develops the classical Ash‘arī account under the influence of Avicenna’s logic and psychology to explain the causes of widely-accepted and seemingly self-evident value judgements. But as in earlier Ash‘arism, he goes no further than to offer a subjectivist metaethics meant only to support an anti-realist stance in order to refute Mu‘tazilī realism, not as the groundwork for an alternative ethical system.²⁰ What is radically new in his account—new, that is, in the context of *kalām* metaethics—is that it recognises two types of ends that give rise to value judgements: conscious ends and unconscious ends. The latter are not ordinarily detectable because of the erroneous judgement of the estimation, which is twofold: it perceives the subjective consequences of things as real and intrinsic to them, and it tends to generalise value judgements, thus extending their scope. This position comes into direct conflict with classical-*kalām* epistemology, as it undermines the principle of immediate knowledge (*‘ilm ḍarūrī*). Much of what appears to the overwhelming majority of sound-minded people to be immediately true may turn out to be falsehood generated by the estimative faculty. For the Mu‘tazila, the notion that God could equip humans with a mental faculty that distorts their perception and understanding of reality would be an evil act on his behalf. While they accept that he creates things both within and outside humans, such as desires, which motivate them to choose to commit bad acts, he cannot create them with minds that could deceive them and at the same time treat them as accountable for their choices. Therefore, the Mu‘tazila, as already noted, recognise only conscious ends: an agent knows that a certain act (say, telling a lie) will benefit her (by producing financial gain) and that the act is intrinsically bad, and then chooses either to perform or not to perform it. If a sound-minded agent uses their mind, they will know if an act is beneficial or harmful, or if it is good or bad, and will not confuse one type of judgement for the other. It was common in earlier *kalām* for theologians to dismiss specific views that their opponents claim to be immediately known as false; al-Ghazālī is the first theologian to offer a robust explanation of their falsehood.

20 This is true of al-Ghazālī’s ethics of action, as set out in his juristic and theological works, but not necessarily of his virtue ethics, where, under philosophical influence, he assigns a normative function to the practical intellect (*al-‘aql al-‘amalī*) (al-GHAZĀLĪ 1964: 203ff.).

3. Al-Malāḥimī: A Failed Defence of Ethical Realism

Al-Malāḥimī's defence of the Mu'tazilī ethical theory against al-Ghazālī's attack occurs in the discussion on "the judgements that apply to acts" (*aḥkām al-af'āl*), that is, the value judgements of ethically evaluable acts (al-MALĀḤIMĪ 2007: 830-54). The discussion is motivated by three main objectives, each treated in one or more dedicated sections—namely, (1) to determine the definitions, or realities (*ḥaqīqa*), of these judgements, (2) to establish that there are in fact acts in existence to which these judgements apply, and (3) to determine the causes of these judgements. We are concerned here only with the last two.

To establish that ethically evaluable acts actually exist—the second objective—is to affirm goodness and badness as actual attributes of acts, as opposed to mere fanciful descriptions thereof, and moreover to affirm them as *real* attributes of acts as objects in the external world. Al-Malāḥimī views the task as essentially effortless; for like earlier Mu'tazila, he maintains that some value judgements are known immediately (*darūrī*). He opens the section titled "Affirming (*ithbāt*) good acts" as follows:

Know that all that is needed to affirm this is to draw attention (*tanbīh*) [to the fact], rather than to infer [it] (*istidlāl*). For every sound-minded person knows that there are some acts on account of which no blame is deserved, such as all that is beneficial,²¹ causes no harm to anyone, and is characterised by none of the configurations of badness. Therefore, [goodness] in general (*'alā l-jumla*) is affirmed and established through reason. (al-MALĀḤIMĪ 2007: 831)

The same point is made in a section titled "Affirming bad acts" (al-MALĀḤIMĪ 2007: 840-1; 845-6).²² Because the existence of bad acts is known immediately, it cannot be inferred from evidence. All that one can do to confirm it is to "draw attention" to the immediate knowledge that we already have of the badness of certain acts. Being self-evident, these value judgements, al-Malāḥimī submits, are agreed upon by all sound-minded people, including Muslims and adherents to other belief systems. And it is impossible for any sound-minded person not to recognise their truth, even if their thinking is misguided by some factor or other, such as a misbelief obtained through specious reasoning or uncritical imitation of others. So long as a mind remains sound, nothing could corrupt its ability to possess items of immediate knowledge where it should have them. Denials of the truth of any such immediately-known ethical facts are dismissed as disingenuous (al-MALĀḤIMĪ 2007: 845-6).

In the tradition of earlier Baṣran Mu'tazilī sources, al-Malāḥimī also refutes the counter-thesis to ethical realism—that ethical value judgements of approval and disapproval ordinarily passed on acts are subjective, and as such of the same order as aesthetic judgements of approval and disapproval passed on visible forms (cf. 'ABD al-JABBĀR 1962: 19-21). This anti-realist, subjectivist thesis is readily implied in the two central Arabic expressions employed to denote ethical values, *ḥasan* (good) and *qabīḥ* (bad), which are widely used to mean, respectively, "beautiful, attractive" and "ugly, repulsive". Al-Malāḥimī contends that the two types of judgement are poles apart (al-MALĀḤIMĪ 2007: 841-2). Ethical judgements

²¹ Reading *kull* for *akl*.

²² Reading *mā lā yajidu* for *mā yajidu* at 841, l. 6.

are rational: they are real, and as such apprehensible to the mind. However, when a person finds a thing either pretty and desirable, or ugly and repulsive, the judgement will arise from the person's self (*nafs*) depending on the extent to which that thing is aligned with her desire (*shahwa*) and repulsion (*nafra*). For this reason, the same thing may be attractive and desirable to one person, and unattractive and undesirable to another person. Such differences betray the subjective nature of these judgements, just as the (purported) agreement of all sound-minded people on certain ethical value judgements confirms their objectivity and truth.

This much was normally sufficient, in earlier Mu'tazilī sources, to eliminate the subjectivist counter-thesis. Because of al-Ghazālī's criticism, however, al-Malāḥimī revisits this counter-thesis in the next section, "That on account of which bad acts are bad," in which the causes of the ethical attributes of acts are determined. Half of the section is devoted to responding to al-Ghazālī, whom al-Malāḥimī does not name, but clearly intends when he refers to "one of the later ones among our adversaries, who had learned some of the teachings of the philosophers" (al-MALĀḤIMĪ 2007: 846). Aside from implying that al-Ghazālī had only a mediocre grasp of philosophy, al-Malāḥimī highlights the philosophical influence on him to score a polemical point, as philosophy was still widely seen as a heterodox system of thought. He then starts by summarising some of the criticisms deployed in the *Mustasfā* against the Mu'tazilī theory of ethical value (al-MALĀḤIMĪ 2007: 846-8).

First, however, he complains that al-Ghazālī misunderstands the Mu'tazilī position. Al-Ghazālī claims that the Mu'tazila are ethical essentialists, as al-Malāḥimī points out:

He reports that our school members hold things that they do not in fact say, believing these to be their teachings—namely, that they hold that what is good and what is bad are good or bad on account of their essence (*li-dhātihī*). This betrays his ignorance of the position of our school members. So, there is no point in reproducing those parts of his discussion. [...] He then argues that good and bad [acts] are not good or bad on account of their essence; but this is not what we actually hold. (al-MALĀḤIMĪ 2007: 846-7; cf. al-GHAZĀLĪ n.d.: I, 178-9; I, 182-3)

Al-Malāḥimī, of course, is correct. In the vein of most earlier Ash'arī treatments of the subject, al-Ghazālī portrays the Mu'tazila as espousing an essentialist theory of ethical value. Claiming that they maintain that badness is an 'essential attribute' (*wasf dhātī*) of the act, he refutes this position simply by adducing acts that are good in some cases, but bad in others—evidence that ethical value is not essential to acts (al-GHAZĀLĪ n.d.: I, 183). It is striking that al-Ghazālī misrepresents the Mu'tazilī view in this manner, considering that in his main work on jurisprudence, the *Burhān*, his teacher al-Juwaynī criticises fellow Ash'arīs and Shāfi'īs for misunderstanding the Mu'tazilī position in precisely this way:

Those who reported the position of [the Mu'tazila] differed on [how to interpret] their statement, 'A thing is bad or good in itself (*li-'aynihi*)'. It has been reported that they hold that the badness and goodness of acts²³ are among their attributes of essence. It has also been reported that badness is an essential attribute, but not goodness, or vice versa according to [Abū 'Alī] al-Jubbā'ī [d. 303/915]. All of this betrays ignorance of

²³ Reading *al-maf'ūlāt* for *al-ma'qūlāt*.

the position of [the Mu‘tazila] (*jahl bi-madhhabihim*). What they actually mean by saying, ‘A thing is bad or good in itself’, is that [the badness or goodness of the act] is apprehended by one’s mind, without needing to be informed [of these judgements] by another. (al-JUWAYNĪ 1979: I, 88-9)

In other words, ‘in itself’ here means ‘intrinsically’, rather than ‘essentially’. Accordingly, in his theological *magnum opus*, the *Shāmil*, al-Juwaynī provides a more accurate account of the Baṣran Mu‘tazilī theory of act-configurations.²⁴ Al-Juwaynī, after all, was closely familiar with Mu‘tazilī sources, particularly ‘Abd al-Jabbār.²⁵ Yet, with the political and intellectual decline of Mu‘tazilism by the late fifth/eleventh century, al-Ghazālī’s attention shifts towards new and more urgent threats, specifically the philosophers and Ismā‘īlīs. And it is for this reason, it seems, that he shows little interest in offering an accurate account of the Mu‘tazilī theory.²⁶

Page | 94

Al-Malāḥimī chooses to ignore al-Ghazālī’s off-target criticisms and instead focuses on the subjectivist lexical definitions he gives for ‘goodness’ as ‘agreement with the agent’s ends’, and ‘badness’ as ‘disagreement with the agent’s ends’, and his claim that value judgements said to constitute immediate knowledge of the external world are often in fact figments that originate in the estimation. In defence of Mu‘tazilī realism, al-Malāḥimī responds by deploying three arguments.²⁷ The first two seek to show that the consequences of acts on their agents cannot account for widely-held value judgements, because the two are not always correlated.

The first argument is that ethical value is not reducible to subjective ends, because some acts are bad but not harmful. Al-Malāḥimī opines that we judge such acts to be bad on account of their intrinsic badness, which is apprehensible to reason, but do not find them disagreeable to our disposition (*tab*). He gives the examples of the mental act of adhering to a misbelief (*jahl*), such as believing that the sky is below us and the earth above us, and purposeless acts (*‘abath*), such as speaking to inanimate objects.²⁸ Al-Malāḥimī invokes acts that are not religiously prohibited to exclude the explanation that their badness originates in the teachings of revelation. He then considers the possible rejoinder that seemingly harmless acts are often in fact not so harmless, because they involve the expenditure of effort, and he counters that even if we postulate that the performance of a purposeless act brings its agent a benefit that

24 Al-JUWAYNĪ 2010: II, 731 ff. In his shorter theological work, the *Irshād* (1950: 257 ff.), he does not discuss act-configurations, but still avoids characterising the Mu‘tazilī theory as essentialist.

25 As is clear from one telling reference he makes to ‘Abd al-Jabbār’s *Mughnī* (SHIHADĒH 2013: 193).

26 Hourani explains al-Ghazālī’s inaccurate account of the Mu‘tazilī theory thus: “The absence of living challengers was taking its toll on the level of argument of Sunnite theologians, as it had done already on that of Ghazālī’s predecessor Juwaynī” (1976: 82; cf. 1975). However, as already noted, al-Juwaynī is in fact careful to provide an accurate account of the Mu‘tazilī position, whereas al-Ghazālī intentionally misrepresents it. The latter’s motives lie in his conception of *kalām* as a pragmatic, dialectical art, a subject that goes beyond the scope of the present study (SHIHADĒH 2005: 142 ff.; 2015).

27 AL-MALĀḤIMĪ 2007: 848-9. He then attacks Ash‘arī divine command ethics (2007: 849-51).

28 AL-MALĀḤIMĪ 2007: 848. On *jahl* in the sense of misbelief, see SHIHADĒH 2013. For a discussion of the Mu‘tazilī position on the badness of purposeless acts, see LEAMAN 1980, although he assumes that ‘Abd al-Jabbār was the first to hold that purposeless acts are bad. This in fact is a standard Mu‘tazilī view, which predates ‘Abd al-Jabbār.

outweighs the minor effort expended, the act would still be bad (al-MALĀḤIMĪ 2007: 848-9). This argument, however, fails two of al-Ghazālī's tests. For, firstly, al-Malāḥimī appears to concede that simply by virtue of being acts, all purposeless acts involve an inconvenience to the agent. So by al-Ghazālī's characterisation of the first type of error, they would be judged bad on subjectivist grounds by their agent, whose estimative faculty may present their badness as a property intrinsic to them. The second type of error would be to assert that purposeless acts are universally bad, even when, as in the case postulated by al-Malāḥimī, an agent occasionally gains a benefit greater than the effort expended in performing such an act. By the same token, one may argue that adhering to a misbelief is often contrary to an agent's ends, because it may result in misguided action

In his second defence, al-Malāḥimī goes a step further, arguing that value judgements often run counter to an agent's self-centred, prudential inclination, because some harmful acts are in fact good and obligatory (al-MALĀḤIMĪ 2007: 849). These include types of just action, such as paying back a debt and treating adversaries fairly in debate, which tend to involve burdensome inconvenience to their agent. If ethical value were correlated to an act's consequences, people would have agreed on the badness of these acts. Again, the argument seems to miss its target, because these acts seem no different than the act of lying, which al-Ghazālī considers under the second type of error. It is arguable that even though wrongful acts are often beneficial to the wrongdoer, most instances of wrongful acts are harmful to the community at large and for this reason are judged bad. This value judgement is accordingly inculcated in individuals.

In both arguments, al-Malāḥimī fails to address and eliminate al-Ghazālī's account of the causation and epistemic status of value judgements. He highlights the philosophical influence evident in his older contemporary's refutation of Mu'tazilī ethics and provides an accurate summary thereof, but nonetheless treats it as a much less sophisticated form of subjectivism than it actually is. Al-Ghazālī proposes that value judgements often derive from unconscious ends, which he explains through the workings of the estimative faculty, whereas al-Malāḥimī, deploying the outdated toolkit of earlier Mu'tazilism, recognises only conscious ends. From a dialectical standpoint, the outcome is a clear win for al-Ghazālī.

Al-Malāḥimī takes a different tack in his third argument, in which he defends the rationality of value judgements. To the view that value judgements derive from the disposition rather than from reason, he responds as follows:

Suppose an act is either entirely harmful or entirely beneficial to a sound-minded person, [1] will he differentiate between the two? And if he does differentiate between them, [2] will reason then dictate to him that he ought to obtain what is entirely beneficial to him, and that if he benefits from that he will not deserve blame from other sound-minded people [...]? As for what is entirely harmful to him—such as striking or injuring his own body, or wasting his wealth for no purpose—will reason dictate to him that he ought not do this, and that [if he were to do it] it would be right for other sound-minded people to say to him, 'Why did you do this!'

If [our opponent] answers that reason dictates none of this to sound-minded people, he will be speaking disingenuously and will be equating the behaviour of sound-minded people who have no knowledge of revealed religions with the behaviour of insane people who cannot differentiate between what they are entitled to do and what

they are not entitled to do. However, if he answers that reason does indeed dictate this, we will say to him: But this is exactly what we mean by ‘rational badness and goodness’ (*al-qubḥ wa-l-ḥusn al-‘aqlī*). So, your claim that reason can make no judgement of goodness or badness on acts before [the reception of] revelation is false. (al-MALĀHIMĪ 2007: 849)

Al-Malāhimī says that the role of reason here is to recognise, first, the ethical value of an act (the *is*) and, second, that it is imperative on the agent to perform it, or to refrain from it (the *ought*). He argues that his adversaries will be forced to concede both of these functions of reason. But his former, ontological claim is problematic. The examples given are acts whose ethical value is ontologically ambiguous, because acts whose only consequence is to benefit their agent, or to harm their agent (such as causing injury to oneself), and hence have no direct impact on other living beings, are recognised by both realists and subjectivists as, respectively, good and bad, although the two sides will differ on the precise referents of ‘good’ and ‘bad’ here and on the grounds of the goodness and badness of these acts. The ethical rationalism of the Mu‘tazila does not consist of asserting the mere fact of the goodness and badness of these and other acts, but also requires the recognition of the reality of goodness and badness as attributes of acts and the causes of these attributes—both of which notions are rejected by al-Malāhimī’s adversaries. So his claim that this is exactly what the Mu‘tazila mean by ‘rational badness and goodness’ is misleading, and seems to present the debate as one between ethical cognitivism and non-cognitivism, rather than a debate between ethical realism and anti-realism. What is more, Ash‘arīs do not deny that the mind is able to differentiate between harmful and beneficial things; they accept that internal sensations are objects of knowledge—introspective knowledge—but deny that these cognitions derive from reason. As for the latter, *ought* claim, al-Malāhimī expects Ash‘arīs to agree that reason provides the agent with the imperative to act or not to act, and others with the entitlement to praise or to blame an agent for performing an act. However, they would simply deny this claim; for although they appeal to subjectivism dialectically in a debate on metaethics, they do not subscribe to a consequentialist normative ethics. Al-Juwaynī maintains that the imposition of obligations (*taklīf*) on agents is not a function of reason, but the prerogative of revelation (al-JUWAYNĪ 1950: 258).

Overall, al-Malāhimī’s defence of ethical realism against al-Ghazālī’s philosophically-influenced criticism shows that Mu‘tazilism at this crucial juncture was unable to keep up with its adversaries. Although the decline of Mu‘tazilism in the fifth/eleventh century was to a great extent the outcome of socio-political circumstances, what we see here is evidence that the school was losing the intellectual battle as well. This observation is, of course, a historical one and should not be taken to imply that the ethical realism taught by the Baṣran Mu‘tazila was inevitably doomed to failure. Whether later, Zaydī Mu‘tazilism manages to catch up, so to speak, and to put forth a more compelling response to neo-Ash‘arī ethical thinking remains an open question, and certainly one worth pursuing.

4. Al-Rāzī: From Subjectivism to Consequentialism

In the discussion on ethical value in his earlier theological work, *Nihāyat al-‘uqūl fī dirāyat al-uṣūl*, Fakhr al-Dīn al-Rāzī offers two lines of refutation of the Mu‘tazilī epistemological claim that the ethical value of certain types of act is self-evident, which correspond to the second and third tasks of al-Malāḥimī. The first is simply to show that purportedly self-evident value judgements are not in fact self-evident (al-RĀZĪ 2015: III, 275-6). The second, “more powerful” line goes a step further by acknowledging the prevalence of certain ethical maxims among people, and then arguing that the nature and grounds of ethical value are not what they are claimed to be in Mu‘tazilī ethical realism (al-RĀZĪ 2015: III, 276-8; SHIHADĒH 2006: Ch. 2). Al-Rāzī briefly proposes three alternative explanations for value judgements. The first two ground value judgement in an individual’s emotive reactions to a thing or occurrence. Some judgements, he first argues, are engendered by the emotive attraction and repulsion that an act arouses in the disposition, such as the judgements that justice is good, and wrongful action bad. Some are engendered by self-centred prudence, such as the judgement that it is good to assist a suffering human or animal, which arises from the pain experienced by a tender-hearted observer. His third explanation grounds some judgements in their consequences for society, but is, in the final analysis, likewise subjectivist, as we shall see shortly: some judgements, he says, are engendered by the consensus of people on rules that guarantee the wellbeing of society, such as the principles that lying and wrongful action are bad. The Mu‘tazila, al-Rāzī argues, fail to eliminate these alternative grounds of value judgements when establishing their own, realist account. He immediately then considers the possible response that these alternative explanations have in fact already been addressed by al-Malāḥimī in the *Mu‘tamad*, and he paraphrases the latter’s first two responses to al-Ghazālī.

Al-Rāzī counters al-Malāḥimī’s first argument by expanding on his emotivist account of value judgement. Al-Malāḥimī had argued that because some acts, such as purposeless acts, are harmless but nonetheless bad, an act’s badness cannot be reduced to an emotive reaction to its actual or expected consequences. This argument, al-Rāzī retorts, can go no further than proving that the badness of (purportedly) harmless acts cannot be explained through their consequences, and thus falls short of establishing the general proposition that the badness of all bad acts cannot be thus explained, for which a further proof is needed (al-RĀZĪ 2015: III, 278-9; 280-1). He contends that harmful action in fact is not the only trigger for emotive, dispositional repulsion (*nafra ṭabī‘iyya*), because the latter is experienced at the perception of things that are entirely harmless to the perceiver, such as people with bodily defects or menial jobs. Indeed, the perceiver may benefit from the menial work of others, yet still experience the same repulsion towards them. It is perfectly conceivable, therefore, for purposeless action and holding a misbelief to be entirely harmless to their observers, and yet arouse a similar emotive repulsion in them. Al-Malāḥimī fails to rule out this possibility when he claims that the repulsion we experience towards bad acts derives from reason (*nafra ‘aqliyya*) rather than disposition, which is to say that it is grounded in our knowledge of the external world, rather than in emotive repulsion.

In his response to al-Malāḥimī’s second argument, al-Rāzī elaborates on his third explanation of value judgements, which grounds them in an act’s consequences for society,

and he argues that it too reduces to subjectivism. Al-Malāḥimī had observed that value judgements often run counter to their subjective consequences, because we judge some acts to be good even when they are harmful to us, or bad even when they are beneficial to us (al-RĀZĪ 2015: III, 279; 281-2). In response, al-Rāzī advances a more sophisticated account of ethical motivation than the one refuted by al-Malāḥimī (cf. SHIHADĒH 2006: 78 ff.). What the latter attacks is a simple subjectivism, according to which the agent judges an act to be good or bad depending on the benefit or harm he expects from it. So, because wrongful action tends to be advantageous to its agent, those who commit such acts should, by this reasoning, recognise them as good; however, all people actually recognise wrongful acts as bad.²⁹ Al-Rāzī takes the view that because agents can only be motivated by self-centred interest, the agent of a wrongful act will commit it only if he believes that it is beneficial and hence subjectively good. What he concedes here is that such an agent would nonetheless accept the general ethical maxim that wrongful action is bad, which seems to run counter to his subjectivism. Al-Ghazālī, as we have seen, explains such value judgements as mental errors arising from the psychological faculty of estimation. Al-Rāzī does not appeal to faculty psychology here, but instead offers an explanation that can be best described as an exercise in social psychology. He argues that value judgements often arise out of a calculus that involves not only the direct consequences of individual acts, but moreover the consequences of the verbal act of assenting to ethical rules. The consequences that an act leads to (*yu'addī ilā*) can be either temporally immediate or anticipated in the future (*ḥālan aw ma'ālan*). Al-Rāzī reasons that the act of assenting to certain ethical rules publicly is likely to result in indirect, future consequences that are favourable to its agent, whereas denying them is likely to result in adverse consequences for its agent by normalising types of action that are harmful to him, and that people's awareness of these anticipated consequences motivates them to assent to those rules. For instance, if one proclaims (*aftā bi-*) that wrongful action is good, this assertion will consequently undermine the ethical rule that one ought to refrain from wrong action and thereby render him susceptible to the wrongful action of others. This subjectively adverse consequence thus motivates people, even those who commit wrongful acts, to accept the badness of wrongful action. And because all sound-minded individuals recognise that they have a stake in such value judgements, society at large will consent (*tawāḍa'ū*) on them, thus giving rise to widely-accepted ethical rules. The same, al-Rāzī remarks, applies to the widely-accepted maxims that filial piety, fairness, justice and keeping promises are good and obligatory. There is evidence elsewhere in his works that this analysis of ethical motivation should serve as the basis of a normative ethics; this would clearly yield a form of rule-consequentialism as opposed to a simple act-consequentialism.

So, al-Rāzī identifies two origins for seemingly non-self-centred value judgements. The first is that some arise out of emotive, dispositional attraction and repulsion, partly out of the dispositional "tenderness associated to the genus" (*al-riqqa al-jinsiyya*, as it occurs in al-Rāzī's works) (SHIHADĒH 2006: 52; 78). This is a rather traditional explanation, which we encounter in al-Ghazālī and earlier Ash'arī sources. The second is that judgements often

²⁹ Reading *yashtahira li-l-zālim* (or *li-l-zalama*) and *yashtahiru li-l-mazlūm* for *tashtahira al-zalama* and *shahara al-mazlūm* (at 279, ll. 7-8).

originate from the self-centred subjective calculus just described. Some judgements—for instance, ‘Wrongful action is bad’—seem to arise out of the combination of both processes.

Al-Rāzī does not address al-Malāḥimī’s third argument—that even a subjectivist must accept a rational conception of value and obligation—which, as noted, does not pose much of a threat from a dialectical perspective. After all, al-Rāzī in the *Nihāyat al-‘uqūl* and later works departs from earlier Ash‘arism by making precisely that same point—that his subjectivist conception of ethical value is a rational one, in that the subjective goodness or badness of acts is knowable through reason. This point is absent in an earlier theological work of his (al-RĀZĪ 2007: 206 ff.), and it is most likely an influence first and foremost from the third argument of al-Malāḥimī, whose works al-Rāzī starts to engage with very closely in the *Nihāya*, particularly in the discussion on ethical value. Al-Rāzī may have been secondarily influenced by a passage in al-Juwaynī’s later juristic work, the *Burhān*, in which he states that reason requires the agent to pursue what is subjectively beneficial and to avoid what is subjectively harmful (al-JUWAYNĪ 1979: I, 91).³⁰ Unlike al-Rāzī, however, al-Juwaynī’s view is not articulated prominently and systematically, and hence had limited impact on later sources.

Thus, in the *Nihāyat al-‘uqūl* and mid-career works, al-Rāzī says that goodness and badness, defined respectively as agreement or disagreement with the disposition, are rational concepts (‘*aqlī*’), in that the agent apprehends the subjective value of an act through reason. Defined, respectively, as ‘not deserving of punishment’ and ‘deserving of punishment’, they are by contrast ‘religious’ (*shar‘ī*), because acts become punishable only through God’s command, which is received through a divinely-revealed religion (al-RĀZĪ 2015: III, 247; cf. SHIHADĒH 2006: 56 ff.).³¹ That, in contrast to earlier Ash‘arīs, al-Rāzī characterises introspective knowledge of emotions as ‘*aqlī*’ is not a trivial shift of usage, but reflects a departure from the classical-Ash‘arī conception of intellect. Whereas ‘intellect’ (‘*aql*’) was earlier defined as a body of immediate knowledge correlated to facts about the external world, al-Rāzī defines it as an innate capacity (*gharīza*) through which knowledge is gained (al-RĀZĪ 1991: 250-1). (The theological and philosophical background of this development goes beyond the scope of the present study and will be investigated in a forthcoming study.) The scope of ‘*aql*’, thus conceived as the capacity for cognition, accordingly encompasses all immediate and non-scriptural inferential knowledge, including introspective knowledge. An agent will therefore be able to apprehend the pain (or pleasure) that an act causes him and accordingly make a value judgement on the act in the form of a proposition—‘That act is bad (or good)’, where ‘bad’ and ‘good’ are defined respectively as a direct or indirect cause of pain, or of pleasure.

This position represents only a minor departure from the position of al-Ghazālī and earlier Ash‘arīs, as the role assigned to reason here is only to recognise the subjective value of an act (the *is*), as opposed to establishing obligations—that is, acts that *ought* to be performed or omitted. It is, nonetheless, an important development in that direction, because in later works al-Rāzī goes further to deny the religious definition of ethical value terms, and to assert that even religious obligations, established through divine command and prohibition, have a

³⁰ As for why I do not consider al-Ghazālī to be a significant influence on this point, see fn. 20 above.

³¹ Reading *nadhhabu* for *dhababa*, at l. 1.

rational basis. It is reason that provides the agent with the obligation to adhere to the divine law in order to avoid severe punishment in the hereafter. He accordingly declares his conception of ethical value and obligation to be a rational one, although, being a consequentialism, it is antithetical to Mu'tazilī realism (al-RĀZĪ 1987: III, 289-90; cf. SHIHADĒH 2006: 63 ff.). This is the first time in Ash'arism that a rational conception of obligation is adopted systematically as the main principle of normative ethics, and that a theory of divine command ethics becomes subsumed within this rational framework.³²

Concluding Remark

To conclude, let us briefly revisit the point made at the start of this article concerning the supposed anti-rationalism of Ash'arī ethics. This characterisation belonged to the old narrative, which depicted the classical period of Islamic thought as a period in which the banner of rationalism was held aloft by philosophy and Mu'tazilism, but was then superseded by anti-rationalist theological currents spearheaded by Ash'arism. We have shown that this characterisation of Ash'arī ethics is little more than a caricature: Ash'arīs were not only ethical rationalists, but moreover, in the late eleventh and twelfth centuries, the more critical and innovative ethical thinkers. This calls on us to reflect upon the way in which often murky and heavily value-laden categories such as 'rationalist' and 'anti-rationalist' are employed more widely in the field.³³

Bibliography / References

Sources

- ‘ABD AL-JABBĀR al-Hamadhānī. *Al-Mughnī fī abwāb al-tawhīd wa-l-‘adl*. Vol. 4: *Ru’yat al-bārī* / ed. Muḥammad M. Ḥilmī and Abū l-Wafā al-Ghinīmī. Cairo 1965.
- . *Al-Mughnī fī abwāb al-tawhīd wa-l-‘adl*. Vol. 6.1: *al-Ta’dīl wa-l-tajwīr* / ed. Maḥmūd M. Qāsim. Cairo 1962.
- al-BĀQILLĀNĪ, Abū Bakr Muḥammad ibn al-Tayyib. *Kitāb al-Tamhīd* / ed. Richard J. McCarthy. Beirut 1957.
- . *Al-Taqrīb wa-l-irshād* / ed. ‘Abd al-Ḥamīd ‘A. Abū Zunayd. 3 vols. Beirut 1998.
- al-GHAZĀLĪ, Abū Ḥāmid. *Al-Iqtisād fī l-i’tiqād* / ed. Ibrāhīm A. Çubukçu and Hüseyin Atay. Ankara 1962.
- . *Mihakk al-nazar*. Cairo 1925.
- . *Mi’yār al-‘ilm* / ed. Sulaymān Dunyā. Cairo 1961.
- . *Mizān al-‘amal* / ed. Sulaymān Dunyā. Cairo 1964.
- . *Al-Munqidh min al-ḍalāl* / ed. Farid Jabre. Beirut 1969.
- . *Al-Mustasfā min ‘ilm al-uṣūl* / ed. Ḥamza Z. Ḥāfīz. 4 vols. Medina n.d.

³² On how this consequentialism is reflected in al-Rāzī’s account of the objectives (*maqāṣid*) of the divine law and *maṣlaḥa*, see SHIHADĒH 2006: 63 ff.

³³ I am grateful to Ferial BOUHAFĀ and Mairaj SYED for their comments on an earlier draft of this paper. All shortcomings are my own.

- IBN FŪRAK, Abū Bakr Muḥammad ibn al-Ḥasan. *Mujarrad maqālāt al-Shaykh Abī l-Ḥasan al-Ash‘arī* / ed. Daniel Gimaret. Beirut 1987.
- IBN MATTAWAYH, al-Ḥasan ibn Aḥmad. *Al-Tadhkira fī aḥkām al-jawāhir wa-l-a‘rāḍ* / ed. Daniel Gimaret. 2 vols. Cairo 2009.
- IBN SĪNĀ, al-Ḥusayn ibn ‘Abdallāh. *Kitāb al-Ishārāt wa-l-tanbīhāt* / ed. J. Forget. Leiden 1892.
- . 1938. *Al-Najāt*. Cairo 1938.
- . *Al-Shifā’, al-Nafs* [Avicenna’s *De Anima (Arabic Text), Being the Psychological Part of Kitāb al-Shifā’*] / ed. Fazlur Rahman. London 1959.
- . *Al-Shifā’, al-Manṭiq, V. al-Burhān* / ed. Abū l-‘Ulā al-‘Affī. Cairo 1956.
- al-JUWAYNĪ, Abū l-Ma‘ālī ‘Abd al-Malik. *Al-Burhān fī uṣūl al-fiqh*. 2 vols. / ed. ‘Abd al-‘Azīm al-Dīb. Doha 1979.
- . *Kitāb al-Irshād* / ed. Muḥammad Mūsā and ‘Alī ‘Abd al-Ḥamīd. Cairo 1950.
- . *Al-Kāmil fī ikhtisār al-Shāmīl* [an abridgement of the *Shāmīl* by a certain Ibn al-Amīr] / ed. Jamāl ‘A. ‘Abd al-Mun‘im. 2 vols. Cairo 2010.
- . *Kitāb al-Talkhīṣ fī uṣūl al-fiqh* / ed. ‘Abdallāh J. al-Nībālī and Shubbayr A. al-‘Umarī. 3 vols. Beirut 1996.
- al-MALĀḤIMĪ, Rukn al-Dīn. *Al-Fā’iq fī uṣūl al-dīn* / ed. Fayṣal B. ‘Awn. Cairo 2016.
- . *Al-Mu‘tamad fī uṣūl al-dīn* / ed. Wilferd Madelung. Tehran 2007. (References are to this edition.)
- . *Al-Mu‘tamad fī uṣūl al-dīn* / ed. Wilferd Madelung and Martin McDermott. London 1991.
- al-RĀZĪ, Fakhr al-Dīn Muḥammad ibn ‘Umar. *Al-Ishāra fī uṣūl al-kalām* / ed. Muḥammad Ṣ. al-‘Āyidī and Rabī Ṣ. al-‘Āyidī. Amman 2007.
- . *Al-Matālib al-‘āliya min al-‘ilm al-ilāhī* / ed. Aḥmad Ḥ. al-Saqqā. 8 vols in 5 (vol. 9 being an independent text). Beirut 1987.
- . *Muḥaṣṣal afkār al-mutaqaddimīn wa-l-muta’akhhirīn min al-ḥukamā’ wa-l-mutakallimīn* / ed. Hüseyin Atay. Cairo 1991.
- . *Nihāyat al-‘uqūl fī dirāyat al-uṣūl* / ed. Sa‘īd ‘A. Fūda. 5 vols. Beirut 2015.

Studies

- AOUAD, Maroun. 1997. “Définition du concept de loué selon le point de vue immédiat dans la Rhétorique du *Šifā’*.” In: HASNAWI et al. (eds.) 1997: 409-51.
- BLACK, Deborah. 1993. “Estimation (*Wahm*) in Avicenna: The Logical and Psychological Dimensions.” *Dialogue: Canadian Philosophical Review*, 32: 219-58.
- . 1990. *Logic and Aristotle’s Rhetoric and Poetics in Medieval Arabic Philosophy*. Leiden: Brill.
- BOUHAFĀ, Feriel. 2021. “The Dialectics of Ethics: Moral Ontology and Epistemology in Islamic Philosophy.” *Journal of Arabic and Islamic Studies*, 21. *IslBh*: 25-54.
- CELLI, Gaia. 2018. “Avicenna and Aristotle’s *Rhetoric*: Lexical Options.” In: WOERTHER (ed.) 2018: 89-115.
- GRIFFEL, Frank. 2012. “Al-Ghazālī’s Use of ‘Original Human Disposition’ (*fiṭra*) and Its Background in the Teachings of al-Fārābī and Avicenna.” *The Muslim World*, 102: 1-32.
- (ed.). 2015. *Al-Ghazālī’s Rationalism and Its Influence. Papers on the Occasion of the 900th Anniversary of His Death in 1111. Vol. 2*. Leiden: Brill.
- HASNAWI, Ahmad et al. (eds.). 1997. *Perspectives arabes et médiévales sur la tradition scientifique et philosophique grecque. Actes du Colloque de la SIHSPAI, 31 mars – 3 avril 1993*. Leuven and Paris: Peeters.

- HOURLANI, George F. 1976. "Ghazālī on the Ethics of Action." *Journal of the American Oriental Society*, 96: 69-88.
- . 1975. "Juwaynī's Criticism of Mu'tazilite Ethics." *Muslim World*, 65: 161-73.
- . 1971. *Islamic Rationalism: The Ethics of 'Abd al-Jabbār*. Oxford: Clarendon Press.
- LEAMAN, Oliver. 1980. "'Abd al-Jabbār and the Concept of Uselessness." *Journal of the History of Ideas*, 41: 129-31.
- MADELUNG, Wilferd. 2015. "Al-Ghazālī's Changing Attitude to Philosophy." In: TAMER (ed.) 2015: 23-34.
- MARMURA, Michael E. 1969. "Ghazālī [*sic.*] on Ethical Premises." *The Philosophical Forum* (New Series), 1: 393-403.
- SCHMIDTKE, Sabine (ed.). 2016. *The Oxford Handbook of Islamic Theology*. Oxford: Oxford University Press.
- SHIHADDEH, Ayman. 2016. "Theories of Ethical Value in *Kalām*: A New Interpretation." In: SCHMIDTKE (ed.) 2016: 384-407.
- . 2015. "Al-Ghazālī and *Kalām*: The Conundrum of His Body-Soul Dualism." In: GRIFFEL (ed.) 2015: 113-41.
- . 2013. "The Argument from Ignorance and Its Critics in Medieval Arabic Thought." *Arabic Sciences and Philosophy*, 23: 171-220.
- . 2006. *The Teleological Ethics of Fakhr al-Dīn al-Rāzī*. Leiden: Brill.
- . 2005. "From al-Ghazālī to al-Rāzī: 6th/12th Century Developments in Muslim Philosophical Theology." *Arabic Sciences and Philosophy*, 15: 141-79.
- SYED, Mairaj U. 2016. *Coercion and Responsibility in Islam: A Study in Ethics and Law*. Oxford: Oxford University Press.
- TAMER, Georges. 2015. *Islam and Rationality: The Impact of al-Ghazālī. Papers Collected on His 900th Anniversary*. Vol. 1. Leiden and Boston, Brill.
- VASALOU, Sophia. 2016. *Ibn Taymiyya's Theological Ethics*. Oxford: Oxford University Press.
- . 2008. *Moral Agents and Their Deserts: The Character of Mu'tazilite Ethics*. Princeton: Princeton University Press.
- WOERTHER, Frederique. 2018. *Commenting on Aristotle's Rhetoric, from Antiquity to the Present*. Leiden and Boston: Brill.