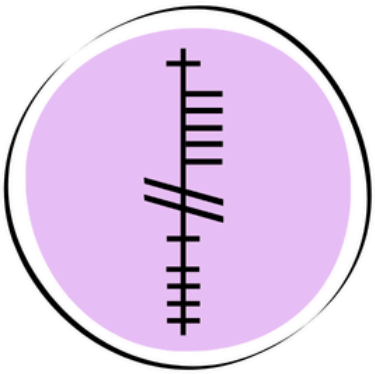
Trinity Student Philosophical Review

Volume IV

Trinity College Dublin 2026



Trinity Student Philosophical Review

Patron

Professor John Divers

Head of the Philosophy Department, Trinity College Dublin

Published in March 2026 by the Trinity Student Philosophical Review,
c/o Department of Philosophy,
Trinity College Dublin, Dublin 2, Ireland

For an online version please visit:
https://www.tcd.ie/philosophy/current-students/trinity-student-philosophical-review/

A catalogue record for this journal is available from Trinity College Library, Dublin.

Printed by printbureau.ie

Cover art by Cate Slattery

The Trinity Student Philosophical Review would like to thank all our sponsors:

Trinity Trust

Tethras Technology Ltd.

printbureau.ie

Trinity Student Philosophical Review Committee - Volume IV

Editor-in-Chief
Shane Burke
burkes34@tcd.ie

Editor
Nina Stremersch
stremern@tcd.ie

Editor
Lara Smith
smithl17@tcd.ie

Editor
Eva Hendley
hendleye@tcd.ie

Treasurer
Tomasz Balcerkiewicz
balcerkt@tcd.ie

Executive Prod. Manager
Giorgia Carli
carlig@tcd.ie

Production & Event Manager
Odhran Killally
killallo@tcd.ie

Social Media Manager
Eva Stylianides
styliane@tcd.ie

Winner of the Iris Murdoch Prize
for the Best Undergraduate Essay 2026

The Genesis of Shame

by James Ryan

Editorial Prelude

I am delighted to welcome you all to Volume IV of the Trinity Student Philosophical Review! The Review was inaugurated with the purpose of highlighting the best work by undergraduate Junior and Senior Sophister students in Philosophy. For the first year since its inception, we have opened up submissions to students from disciplines beyond Philosophy, in recognition of the passion for Philosophy that extends throughout the entire student body. The seven pieces painstakingly chosen for publication by the Editorial Team represent work that goes beyond what would be expected of students at the undergraduate level. We strove for quality of argumentation above all else in this decision, with a secondary view to the overall diversity of topics represented. The pieces we selected all demonstrated creativity and prompted excitement and lively discussion among the Editorial Team. Our passionate agreement and sometimes disagreement over the arguments made was symptomatic of the degree to which the pieces were both thought-provoking and important. We hope that you will share in our passion and excitement.

Philosophy is the only discipline wherein the subject matter and method are one and the same – thought. The identity of content and form in Philosophy entails that nothing remains merely given. Philosophy, unlike the sciences or other humanities, does not presuppose its content as something given from outside of it, to which an equally given method of investigation, with its categories and concepts, is applied. Instead, the thought that is investigated in Philosophy is one and the same thought doing the investigating. This self-relating activity is free in a way that other activities are not, for both the content and form of philosophy emerge simultaneously from the activity. So long as we take ourselves to be free, so long as freedom is grasped as the highest virtue, then Philosophy is the proper vocation of humankind. Such freedom is not merely the *license* of the individual to twist and contort any subject matter they see fit, but is instead the freedom to watch the proper content of Philosophy unfold *itself* spontaneously. That is, Philosophy is not about the vanity of the individual who uses his subjective whim to "subvert and bring to naught everything firm and true", to quote Hegel. Philosophy instead returns to the subject matter itself and allows it to express itself, this subject matter being in truth only thought!

Such a high bar for what constitutes Philosophy seems nearly impossible to meet. However, I believe that each piece in this Volume, in its own way, has

shown this freedom characteristic of Philosophy. I certainly had the impression while reading that the subject matter of Philosophy – thought as such – was showing itself to me, even where the claims of the authors may be narrow and focused. I hope that you all will be struck in the same way as I was by the free thought developed through and within these pieces and will share with me the view that each provides a partial view upon one and the same, ultimate, free thought.

To begin my brief discussion of the content of the piece in this Volume, I am delighted to announce that the **Iris Murdoch Prize** for the best piece goes to **James Ryan** for his paper *The Genesis of Shame*! Ryan's paper masterfully adjudicates between two accounts of shame, the "Adverse Self-Evaluation" model and the "Mere Exposure" model, arguing for a third position that can account for both private and public shame. This model claims that the recognitive rupture between the self and the generalised (ideal) other provides the best philosophical account of shame.

There is perhaps no greater generalised other than God Himself, and what a shame it would be if God rendered all of our choices unfree. Thankfully, **Maisie Norton** argues that God's Grace opens up the possibility of human freedom, acting as a guiding hand rather than a coercive one. Norton makes clever use of the unidirectionality of time to claim that God's omniscience does not render human freedom impossible.

"Every S is P" is about as pervasive in logic as God is omnipresent, but do we know what this categorical form means? **Robert Finan** makes a compelling case for the ambiguity of meaning and consequent nonsense of this categorical form, given that it can be interpreted as two different propositional forms. Instead of throwing the baby out with the bath water, Finan claims that we can keep the conclusions of both classical and contemporary logicians as long as we recognise that they are talking about two different categorical forms.

On the theme of ambiguity and equivocation, **Patrick Byrne** provides a technically rich but accessible defence of expressivism, the view that moral language expresses an evaluative attitude, against one instance of the "Frege-Geach Problem". Byrne shows both that the charge of equivocation against expressivism would also apply to cognitivism (provoking a both or neither dilemma) and that the charge does not apply.

On the topic of morality, **Jessica Kiehstaller** leads us through the thorny and deceptively dense issue of relativism versus objectivism. Kiehstaller demonstrates that the existence of moral disagreements need not automatically support moral relativism, arguing that moral facts may *themselves* be indeterminate in borderline cases, rather than there simply being no moral facts at all.

To complete this moral triumvirate, **Gavin Dunphy** explores the question of whether we have a moral obligation to be polite. Dunphy inserts himself into the debate and dismantles, through pointing out circularity and the effective use of examples, the idea that we have a moral obligation to be polite. Dunphy claims instead that politeness is a tool that we often have an ethical obligation to use.

Finally, **Miriam Treitinger** encourages us to imagine the role of imagination in our disinterested pleasure in aesthetic appreciation. Treitinger's exegesis of Kant enables us to see that we may have disinterested pleasure of non-natural objects as a result of the image-producing faculty of the mind.

As Editor-in-Chief, it has been a privilege and a great pleasure to critically engage with such fantastic work, both that of the published authors aforementioned and those whose pieces did not make it on this Volume of the Review. Each piece represented a singular part, of a plurality of faces, of the one and only concern of Philosophy: thought. This idea was captured wonderfully in the artwork by Cate Slattery, to whom we are greatly indebted.

This fourth Volume of the TSPR could not have been possible without the effort of so many people. It takes a village to raise a child and in much the same way, to bring a publication to print! The entire team of the TSPR has been working incessantly since last October editing, organising, planning, designing and liaising. Without this voluntary labour, none of this would be possible. The support and assistance of the Philosophy Department and especially our Patron John Divers has been invaluable. Particularly, I would like to thank Katy Armstrong and Carly Forde for being so forthcoming in their help and for putting up with an endless barrage of emails! Without their crucial help, we would never have received enough papers to publish, let alone plan a Launch Party and organise a guest speaker. I would like to thank our sponsors, Trinity Trust, Tethras Technology, and printbureau for their overwhelming generosity.

Last, but certainly not least, I would like to thank all of the authors who contributed to our journal this year. You should all be immensely proud of the quality of work and argumentation of which you are capable.

Shane Burke

Editor-in-Chief of the Trinity Student Philosophical Review, Volume IV

Table of Contents

The Genesis of Shame

by James Ryan

Have you no sense of shame? Shame on you! That is why James Ryan has come to us with an astute piece of argumentation: to give us our sense of shame back. When are we ashamed? Is it when we feel *exposed* to another? Or when we feel like we have *failed* ourselves by violating certain ideals that we hold ourselves to? Ryan *exposes* what both these views have *failed* to explain; and thereby puts both to shame. He synthesises a bonanza of Kantian philosophy and Russian literature to argue that shame is best understood as an unequal relationship between oneself and a generalised Other—a relationship which is parasitic to one's very own human dignity.

This essay argues that shame involves the experience of minds in the form of a generalised other, an internalised standpoint of recognition, rather than a literal observer, where shame is a rupture in this recognition of human dignity. Privately, the subject is conscious of the dissolution of the ethic by which she has attained recognition; publicly, a particular person or group embodies an unequal source of evaluation. First, this essay will clarify what it means to experience other minds. Then, it will overview and criticise both the Adverse Self-Evaluation model of shame and the Mere Exposure of shame, respectively, before elaborating this account.

To begin, it is necessary to clarify what constitutes an experience of other minds. To experience other minds does not mean that we experience a given situation from a subjective viewpoint which we distinguish from our own, for that would by its nature be inaccessible to us (Nagel 1986:17). Rather, it is a particular faculty of cognition by which we recognise that we can be perceived by others. More exactly, O'Brien (2011) refers to ordinary self-consciousness as a recognition that we are an object for a generalised other, which in her view entails being seen in a negative light (105). This essay develops this conception in two ways: (1) by arguing that shame does not necessitate being seen in a negative light, and (2) by positing that shame does not need an immediate relation to an evaluator (refuting a claim O'Brien makes in her Social Diminution model). (2020:557) The generalised other is a socially mediated self-consciousness, which is to say, the social process is logically prior to the individual's self-concept (Mead 1934:223), forming the narrative structures through which the agent defines herself. Sartre's concept of 'the Look' is a special activation of this awareness, where the subject perceives herself as the object of another's subjectivity. However, Mead's understanding of the

of the generalised other entails that, even when alone, the mental properties described in 'the Look' continue to make the experience of other minds a simulated reality for the subject.

To further clarify the experience of other minds, it is important here that our conception of other minds is often developed through *analogical* reasoning, inferring mental properties from similarities to our internal mental life. This process takes place in a weaker form for group entities such as when, for example, agents in moral reasoning imagine the cumulative suffering of a large number of people. Empirical evidence demonstrates that people attribute mental properties to groups, largely as a way to reduce the cognitive strain of doing so for separate group-members (Waytz & Young 2012:83). This process takes place even with the deceased: in Homer, Nestor appeals to shame by "asking the warriors" to remember their loved ones "whether living or dead" (Williams 1993:79). Furthermore, it is not merely that one is an object in another's perception that causes shame, as the Mere Exposure model suggests, but specifically that one can be recognised either as an end-in-itself, with dignity, or as a means for utility. Kant ([1785] 2010) holds that all humans must be treated as ends-in-themselves, because this is the only way they can have *unconditional* worth (116). The recognition between the self and the generalised other is the precondition of the worthiness which can prevent shame. This is why people are called 'shameless' not merely when they treat others as a means, but when they also treat themselves as a means. Thus, shame involves an internalisation of one's recognition through the minds of others that does not always literally correspond to how one is perceived, but rather how that perception is mediated through a social understanding of human dignity.

One model of shame does not rely on the experience of other minds. The Adverse Self-Evaluation model of shame argues that shame occurs when an agent construes herself as having failed internalised ethical ideals or a code of conduct (O'Brien 2020:546). This accounts for an important feature of shame: it can be experienced when alone (Zahavi 2014:216). Certainly, shame often includes negative self-evaluation, such as Tolstoy's Rostov: "he regarded himself as a worthless scoundrel whose life could not atone for his crime" (O'Brien 2020:545). The main distinction between guilt and shame in this model is that guilt is over a certain action while shame, such as with Rostov here, concerns the whole self. On this view, while an agent may have developed her ideals within a social context, shame is not specifically a social emotion. In Williams' (1993) discussion of Ajax, Ajax feels shame because he has internalised a certain world whose structure he has violated, which points towards who he is (92). He also rightly notes that *internalisation* is fundamental for there to be such

a thing as shame culture, or to grasp why Achilles would feel shame even if assured he could receive the treasure he was denied by Agamemnon (1993:81). Shame here, then, depends not on the scorn of others but the agent construing herself as 'bad' or 'dishonourable'. This model, then, explains shame in isolation as well as the typically negative character of shame.

However, there are many features the Adverse Self-Evaluation model of shame fails to explain, and it errs when it supposes that this internalisation exclusively involves an ethical code and not a form of recognition. It does not account for the shame that, for example, victims of physical or sexual abuse frequently report. This is a noteworthy counterexample because it does not presume any ethical failing on the part of the victims. Some might respond that these are pathological instances brought upon by a traumatic disruption of ordinary cognition and therefore should not figure in our account of shame. However, there are less severe examples, such as the shame a prisoner experiences being strip-searched, regardless of the prisoner's regard for the authorities. O'Brien (2020) employs the typical example of someone who feels shame due to their minoritised racial or ethnic characteristics, noting that the important factor which the Adverse Self-Evaluation model leaves out is that this shame (1) is a rational statement about facts of the world and (2) involves recognition that the subject has not ethically failed (550). If, however, shame implies an inequality of recognition—where the subject feels herself to be treated by another as merely a means to physical or sexual gratification, or as an object to be used rather than a person to be treated in a certain way—that would explain why shame frequently occurs among people who cannot be faulted, and often do not fault themselves, for any moral failing.

The next main model is the Mere Exposure model, according to which the genus of shame is merely the subject 'construing themselves as exposed to another' (O'Brien 2020:553), without necessitating a concomitant negative self-evaluation. The Mere Exposure model, similarly, addresses features of shame which are inadequately comprehended by the Adverse Self-Evaluation model. For example, it better comprehends shame of nakedness or sexual shame. A conundrum for Augustine was how sexual shame should be understood when sexual intercourse is the only natural means to bring about children (Velleman 2001:32). Velleman (2001), in his reading of the Biblical Fall, views shame as underlying our inability to conceal what we desire to be private. This is why, for example, male frontal nudity is less tolerated, because unwanted erections make men unable to hide their private impulses (39). Likewise, he argues that in the case of the pillory as a Puritan punishment, it was successful at shaming because it 'publicly stripped [the wrongdoer] of his social status as a

self-presenting person' (48). Indeed, it is hard to understand why, on the Adverse Self-Evaluation model, there would be anything more shameful in this case than for the wrongdoer to simply walk around and meet public scorn. The Mere Exposure model, then better incorporates the private/public divide's role in shame, though it does not address private shame very well.

There are many problems with Velleman's account and this model in particular, primarily that it does not provide clear conditions under which exposure results in shame. There are certainly cases of nudity which cause shame that do not have to do with the inability to mask private intentions: for example, being naked before a police officer or military official. There are, conversely, instances where the opposite is the case, such as a nude model posing for art students. O'Brien (2020) brings up the counterexample of the face, which is difficult to control and can sometimes betray far more than genitalia can (555). She also points out that managing shame is not essentially about keeping these things private as much as possible but making one's self-concept more congruent with the recognition from other minds (556-7). A licentious priest, for example, is likely to feel far more ashamed while preaching before a congregation even if his secrets are well-kept.

Furthermore, this model fails by positing a mismatch between the private and public aspects of a person as the cause of shame as such, without identifying what makes these mismatches salient. Take the paradigmatic example of a man who is unfaithful to his wife by seeing a prostitute, which Velleman (2001) refers to as a shameful appearance in public due to the exclusively sexual undertone which would underly it (51). This example, however, demonstrates the failure of both models. With regards to the Adverse Self-Evaluation model, we can concede that guilt might say, "I have gravely violated my own ethical code as a married man," but the more expected response of shame would say something like, "Gosh, imagine what my wife would think if she could see me right now," expressing the virtual presence of other minds within shame. The Mere Exposure model would suggest that the man would feel ashamed both before his wife and before the prostitute. The reason there is a salient difference is that, vis-à-vis the prostitute, they have both recognised each other merely as means to another end, while, vis-à-vis his wife, he has through this act treated her as a means whereas she has treated him as an end. Moreover, the mere act of concealment, and revelation of his private life, is insufficient to induce shame. Perhaps he conceals a childish hobby; if it were discovered, he would feel embarrassed, but he would not feel ashamed. Embarrassment, like guilt, is more concerned with the specific action with relation to others, but not the self-concept of the subject in the minds of others (Zahavi 2014:219). The

Mere Exposure model, however, fails to account for these distinctions.

Now, this essay will more fully develop its account of recognition and its relevance to our experience of other minds within shame. It identifies shame as the particular form of psychosocial stress which occurs when the self and generalised other have engaged in a ruptured or distorted form of recognition. Earlier, this involved shame at treating the other person as a means. This idea illuminates the example of Tolstoy's Rostov, who felt intense shame admitting he wasted his father's money gambling. What O'Brien's (2020) account of this misses is the crucial detail that it is only when the father immediately relents that the son's shame becomes so intense. For Rostov has treated his father as a means but was received by the father as an end: he has figured in the relationship as a different sort of moral agent, and thus it is natural that when the father forgives him that the son should feel this shame so fervently. In an ethical sense, the salient feature is not merely that he has violated a moral code (which would be sufficient for guilt) but that he has disrupted this recognition. This is a similar mechanism to Sartre's (2003) description of shame as 'recognising myself in this degraded, fixed, and dependent being which I am for the Other' (312). In this case, however, the Other figures as the degraded and dependent being. Sartre's description elucidates that the experience of shame by marginalised people is not primarily about the inability to prevent exposure, or ethical self-evaluation, but rather the corrupted form of recognition they receive in racialised society, where the political culture does not treat them as an end with human dignity. This is why people can feel shame for going above the ethical call of duty if they have been violated in some way. For example, a parent may experience shame for making great sacrifices for a child who has become morally depraved, while not believing that it was a mistake to make those sacrifices.

Shame is detrimental to the form of recognition that the agent has developed, which Kant referred to as "a human being's consciousness of his own nobility" (Thomason 2013:229). Kant aids us in understanding sexual shame. He suggested that "sexual desire carries, in itself, a tendency to this kind of solipsism" because it treats people as the object of a sexual appetite (Langton 1995:153), solipsism referring here not to the denial of other minds as such, but the radical reduction of all other entities to a mere means. Kant's characterisation is too extreme, but the frequency in which people are treated as a means within sexual intercourse helps explain why it is such a pertinent source of shame. Augustine's conception of sexual shame as a lack of control can be read here as an incongruence between the actions of the agent in sexual intercourse and their ethical obligations towards the other person. In a private

context, an agent can simulate the mind of the other and in this way experiences her own recognition. This is why the most common invocation of shame occurs when people think to themselves what others would think if they could see the situation. In a public context, the other person embodies this generalised other mind and thus shame strikes with greater force and vivacity. It is also why, in the case of Rostov, the presence of the father induces shame – not because it would not be wrong to waste someone else's money, but because he embodies in his person the perfected example of that recognition which Rostov has corrupted.

People also do not feel that just anyone is capable of shaming them, at least to the same extent. It is difficult for someone who figures as a madman in an agent's mind to induce shame in the agent. Goffman (1956) also outlines the example of a performer who feels ashamed in the presence of an old trainer, who remembers how clumsy and embarrassing the performer once was (101). The performer has no cause to be ashamed by a spectator now, nor did they have cause to be ashamed (even if embarrassed) before the trainer before, but the disruption of recognition incites shame. The trainer embodies the Other insofar as the Other threatens to swallow the subject's self-concept. Similarly, within the examples of physical and sexual violence earlier, one of the aims of the perpetrators in those cases is to diminish the victim's feeling of personal autonomy, which was the precondition to worthy recognition. Thus, this theory of the experience of other minds in shame provides greater clarity for the examples discussed.

In conclusion, shame is fundamentally an experience of the minds of others insofar as it involves a rupture of the self's identity with the simulated self-recognition it ascribes to the generalised other, but it is not dependent on an actual observer or evaluator. It does not presuppose negative self-evaluation, nor is mere exposure sufficient to activate it. Rather, it consists of an inequality of recognition, primarily over whether the self and the generalised other view each other as a means or as an end.

Bibliography

Goffman, Erving. (1956) *The Presentation of Self in Everyday Life*. Woodstock, NY: Overlook Press.

Kant, Immanuel. (2010) "Excerpts from Kant's Groundwork of the Metaphysics of Morals." In *The Ethical Life: Fundamental Readings in Ethical and Moral Problems*, edited by R. Schafer-Landau. Oxford: Oxford University Press.

Langton, Rae. (1995) "Sexual Solipsism." *Philosophical Topics* 23 (2): 149–87. https://doi.org/10.5840/philtopics199523210.

Mead, George Herbert. (1934) *Mind, Self, and Society*. Chicago: The University of Chicago Press.

Nagel, Thomas. (1986) *The View from Nowhere*. Oxford: Oxford University Press.

O'Brien, Lucy. (2011) "Ordinary Self-Consciousness." In *Consciousness and the Self*, edited by JeeLoo Liu and John Perry, 101–22. Cambridge: Cambridge University Press.

———. (2020) "Shameful Self-Consciousness." *European Journal of Philosophy* 28 (3): 545-66. https://doi.org/10.1111/ejop.12596.

Sartre, Jean-Paul. (2003) *Being and Nothingness*. Abington: Routledge.

Thomason, Krista K. (2013) "Shame and Contempt in Kant's Moral Theory." *Kantian Review* 18 (2): 221–40. https://doi.org/10.1017/s136941541300006x.

Velleman, J. David. (2001) "The Genesis of Shame." *Philosophy & Public Affairs* 30 (1): 27-52. https://doi.org/10.1111/j.1088-4963.2001.00027.x.

Waytz, Adam, and Liane Young. (2011) "The Group-Member Mind Trade-Off." *Psychological Science* 23 (1): 77–85. https://doi.org/10.1177/0956797611423546.

Williams, Bernard. (1993) "Shame and Autonomy." In *Shame and Necessity*, 75 102.Berkeley: University of California Press.

Zahavi, Dan. (2014) "Shame." In *Self and Other: Exploring Subjectivity, Empathy, and Shame*, 208–40. Oxford: Oxford University Press.

Is Belief in God Compatible with Belief in Human Freedom?

by Maisie Norton

The question of the compatibility of belief in God with belief in free-will is time old, but Maisie Norton's delightful essay makes it fresh. Approaching the problem of free-will and God's omniscience from the nature of human finitude, Norton makes a compelling case for their compatibility. While God may have caused our existence, we are free to respond to his influence. Most interesting, though, is Norton's argument for the compatibility of God's omniscience with free-will from the nature of temporal existence: God knowing what we will do is no more a limit on human freedom than is the necessity that *something* happens at all.

Belief in God is compatible with belief in human freedom, if only in a non-absolute sense. Specifically, St. Thomas Aquinas's conception of God as the "source of the real life and destiny for men and women now and in the future" (O'Meara, 1997, p.16) is compatible with his understanding of "free-will" as "nothing else but the power of choice" (Aquinas, 1971, I:83:1). Aquinas holds that "forasmuch as man is rational it is necessary that man have a free-will," but that this "free-will is not sufficient thereto unless it be moved and helped by God" (Aquinas, 1971, I:83:1). This assertion is dependent upon Aquinas's account of Grace, and it is problematised by God's omniscience. If God, as Aquinas maintains, knows all things eternally and completely, it would seem that the future is fixed, rendering genuine human freedom impossible. In what follows, I will consider Aquinas's solution to the issue of God's omniscience – namely, that God's eternal mode of knowing does not impose necessity onto contingent human acts. I will establish that a fully theistic metaphysics, such as Aquinas's, is no more inhibiting to an absolutely free will than an atheistic account; and that Aquinas preserves a meaningful conception of human freedom.

If a "free" will is one that is "absolutely independent of any cause" (Dilman, 1999, p.165), then "free-will" cannot be reconciled with belief in the God of Aquinas. Rudi Te Velde summarises Aquinas's conception of the relationship between God and creation as such: "There is no being or acting of any creature unless as caused by God. God is not to be viewed as a particular agent who cooperates with the human agent" (Te Velde, 2006, p.149). However, as outlined by Te Velde, humans' dependence on God does not deprive them of the ability to make choices in line with their own desires, "for God works in each according to its nature" (Aquinas, 1971, I:105:5). Aquinas makes use of the Christian doctrine of Grace, meaning God's unconditional and transformative

love for creation. Aquinas's theology of Grace supposes that free-will, as bestowed upon us by God, is helped by God's Grace, as it "opens human freedom towards God" as a form of "divine assistance (*auxilium gratiae)*"(Te Velde, 2006, p.149). Grace is an "intrinsic (thus created) form of the human agent, by which he can freely and spontaneously be himself" (Te Velde, 2006, p.150). In this way, through Grace, God is "totally and actively present *in a divine manner,* that is without suppressing human freedom and without taking over the role of the human agent"(Te Velde, 2006, p.150).

Grace is thus a kind of offering from God to humans: it is a guiding hand towards beatific realization that one can either take or turn away from by one's own free-will, but the hand will not cease to offer itself. Similarly, a child, not yet endowed with mature judgement or learned experiences, may listen or not to the higher rationality of her guardians when deciding how to cast her will. Unbeknownst to the child, her decisions, her willingness or not to heed her elders, those who love her, will have been influenced by their values and ways of being. By their authority, the guardians, usually parents, can—in most cases—stop the child from doing something against her interests. If they choose not to impose their authority on her, they are allowing her, despite their power to do otherwise, to exercise her freedom. The parents or guardians might hope, however, that by their raising of her, the child has some of their own "good sense" embedded into her character. She can then choose whether or not to act in line with whatever judgement she has inherited, but – and this is crucial by analogy to an Aquinian understanding of free will – she will not be able to remove it from herself. In these earthly terms, taking no God into account, this child would still not be capable of exercising a will that is "absolutely independent of any cause," for she cannot avoid the influences of her parents on her will. She cannot unilaterally divest herself of her acquired intuition. Yet, she still has free-will of a kind: she has the freedom to act in accordance with or defiance of the unavoidable influence of her parents.[1] Likewise, for Aquinas, God's Grace is ever present in us, and God allows us the free-will to relate to that Grace as we so wish or please.

As we have seen, then, for Aquinas, "although the world totally depends on God it has what may be called a 'relative independence' or 'derived autonomy'"(Owens, 1971, p.9). This "derived autonomy," or any sense of human autonomy at all, is, if not threatened, at least potentially foreclosed upon by God's foreknowledge. For Aquinas, God as First Cause created everything *ex nihilo* (out of nothing), and God is eternal. If God is eternal, then God must know everything that is to happen from that eternity, for if God does not, then God "learns things about the world" as they happen, and thus "would be

affected by the world" (McCosker, 2016, p.113). If God could be affected and surprised by the unravelling of worldly events, then God could not be the source of all that currently exists and all that will exist. The eternity of God necessarily means that, in order to be God, God must "know all things intuitively and completely," and God must also know "the nature of everything" as they "pre-exist as an idea in his mind" (Owens, 1971, p.30). Herbert McCabe, in the *Cambridge Companion to the Summa Theologiae,* identifies the thorny question prompted by God's necessary foreknowledge as follows: if "God *does* know from eternity (and still more if from eternity he brings about everything that happens), then it would seem that everything happens determinately," and, therefore, there cannot be "any genuinely free and spontaneous human actions, there cannot be what are known as 'future contingents'" (McCosker, 2016, p.113).[2]

Aquinas deals with this apparent incompatibility between God's foreknowledge and human freedom in his assertion that God's knowing of future events does not necessitate those events. Rather, God, in knowing future events, is merely taking account of the decisions that we *will have* made by our own free-will. Further, there is no sense in using the term "future" regarding God, as God "does not belong within a time-series" (McCosker, 2016, p.113). Aquinas's God "stands at the summit of eternity where everything exists together, looking down in a glance at the whole course of time" (McCosker, 2016, p.115). God sees our own decisions as we make them––not before we make them, but before we know we will have made them. Events "whence to us they cannot be certain, forasmuch as we know future contingent things as such; but (they are certain) to God alone, whose understanding is in eternity above time"(Aquinas, 1971, I:83:2). If everything from across time is contemporary to God, then God's 'foreknowing' is actually just God's "knowing"; and just as we do not necessitate events by knowing them as they happen, nor does God.

Like the child who cannot exercise a will "absolutely independent of any cause," whether there is a foreknowing God or not, everything is still in some way pre-determined by the limits of time and space. Assuming there exists only one reality (as opposed to several or infinite alternate realities), the history of events has therefore unfolded in only one way, and so too will all events henceforth. Today, I do not know what will happen tomorrow. By tomorrow night, I will know what happened earlier that day. Tomorrow will unfold in one way or another, and tomorrow I will know in what way that unfolding will have occurred. It will likely unfold, for the most part, in line with however I choose to conduct myself. And while I am in the process of conducting myself, it will feel to me as though there are alternative ways I might go about my business.

In other words, I am conscious that today I can impact the events of tomorrow: for example, I can stay up very late, and feel tired tomorrow; or I could go to sleep early, and wake up with energy. Nevertheless, by tomorrow night, only one version of events will have occurred—I will have been tired, or not. Right now, I do not know how tomorrow looks, but in a sense—a kind of future perfect sense—it already looks a certain way. Though it seems open ended now, the future already 'is.' It is waiting for its realization as we make decisions. Free will exists in a practical sense—humans make choices that have been influenced by external forces, and future events are contingent on those choices. But when someone makes a single choice out of apparently many alternative options, they were always going to have made that one choice, for the passage of time can only hold in it one version of the past, present, and future. The passage of time yields only one trajectory. Whether there is an omniscient God or not, the future is going to unfold in one way, or in one other way. Aquinas's conception of God, whose foreknowledge merely accounts for the decisions we will have made by our own free-will, no more impinges on human freedom than the future's unidirectional unfolding does.

As such, Aquinas's account of free-will as "rational appetite"—his notion that "one acts with free will when one's deliberation is sensitive to one's own judgements concerning what is best in those circumstances" (Fischer, 2005, p.123) —is compatible with, inasmuch as it is dependent on, an omniscient and omnipotent God, who, through Grace, exists within us, and who is the First Cause and Mover of all acts and actions. Regardless of whether this God exists, there is no such thing as an absolute free-will, for in one way or another "the whole course of a man's life, in all its incidents great and small, is as necessarily pre-determined as the course of a clock" (Schopenhauer, 1951, p.48). This statement is true, not because God as "a particular agent" dominates over the "human agent," but because, like the course of a clock, the future can only account for one trajectory. St. Thomas Aquinas's conception of God is compatible with a belief in human freedom as derived from the freest being.

NOTES

[1] With this analogy, I do not suggest that God is to humans as parents are to their infants. Rather, I aim to (a) illustrate that absolute free-will is inconceivable with or without God, and (b) compare God's Grace to the guiding love of a parent for their child.

[2] A future contingent is an event that may or may not happen, depending on the unfolding over other events.

Bibliography

Aquinas, T. (1971) *Summa Theologiae*. Blackfriars ; London: Eyre and Spottiswoode.

Fischer, J.M. (2005) *Free Will: Critical Concepts in Philosophy*. Routledge.

Ilham, D. (1999) *Free Will: An Historical and Philosophical introduction*. Routledge.

Knight, K. (2017) 'The Summa Theologiæ of St. Thomas Aquinas', *New Advent*. Available at: https://www.newadvent.org/summa/

McCosker, P. and Turner, D. (eds.) (2016) *The Cambridge Companion to the Summa Theologiae*. Cambridge University Press.

O'Meara, T.F. (1997) *Thomas Aquinas theologian*. University of Notre Dame Press.

Owens, H.P. (1971) *Concepts of Deity*. Macmillan.

Schopenhauer, A. (1951) 'Free Will and Fatalism', in *On Human Nature, Essays*. Allen and Unwin.

Velde, R.T. (2006) *Aquinas on God: The 'Divine Science' of the Summa Theologiae*. Ashgate.

The Misguided Debate Over Existential Import: Ambiguity and Aristotle's Categorical Forms

by Robert Finan

Have you ever found yourself stuck in a logic class wondering if the seemingly endless debates over the correct interpretation of logical symbols might be... well... somehow missing the mark? If so, you're not alone! Robert Finan argues convincingly in this essay that the age-old debate between modern and classical logicians over the correct interpretation of Aristotle's categorical forms is fundamentally misguided. Both views, he argues, make the same crucial error: assuming that there is any single correct way to represent the ordinary language expression 'All S is P' in logical notation, because it is, by itself, ambiguous. Fans of formal logic need not worry, though, as Finan concludes that both interpretations remain valuable insofar as they each capture a legitimate meaning of their ordinary language counterpart – they just aren't exhaustive.

Introduction

The history of logic is broadly divided into the study of classical, Aristotelian logic, and the study of contemporary, Fregean logic. The dissonance between these two general approaches has led to many debates. One such debate concerns the nature of Aristotle's traditional categorical forms (i.e. A, E, I, O), specifically with respect to existential import. In this essay, I critically examine the debate between the two primary approaches to existential import, the first being that affirmative sentences have existential import while negative sentences do not (the Classical View) and the second being that particular sentences have existential import while universal sentences do not (the Contemporary View). I ultimately argue that neither of these approaches is preferable, as both make the mistaken assumption that there can be a uniquely accurate, general method for assigning a propositional form to ordinary language expressions of the categorical form **A**. Furthermore, I argue that we ought not discard the surrounding ideas and conclusions derived from either approach, as both can be preserved as useful logical tools.

§1. Terminology and Logical Framework

I begin by outlining some relevant terminology:[1]

- **Fact:** A way that objects in the world are, and are related to one another; a true proposition.
- **Proposition:** A possible fact; that is, the *meaning* of a propositional sign, which throughout this essay will be written in Quantificational Logic (QL, henceforth).

    - **Propositional Sign:** An expression which represents a proposition. A propositional sign *means* the proposition that it represents.
    - **Nonsense:** An expression that is *not* a propositional sign. E.g., 'Purple is equal'.[2]
- **Expression:** A collection of terms in ordinary language (OL, henceforth)—spoken, written, or otherwise manifest—that is either a propositional sign, or is nonsense.
- **Existential Import:** An expression has existential import iff it corresponds to a proposition that states the existence of the expression's subject (S), i.e., a proposition expressible in QL containing (a non-negated instance of) ∃x[Sx].

Logic is primarily concerned with propositions. OL expressions that are non-propositional (e.g., "Hello", "Open the door", etc.) instead fall under the purview of the philosophy of language. That is, logic cannot *directly* deal with nonsense, as these expressions have no corresponding propositions. For the purposes of this essay, all relevant propositions can be written in QL; expressions that cannot be represented in QL will be considered to be nonsense.[3] Throughout this essay, OL broadly refers to all conversational language. 'Tá duine ard amháin ar a laghad ann' and 'There is at least one tall person' both represent,in OL, the same proposition: ∃x[Tx & Px]. Nevertheless, English will be the primary example of OL throughout this essay.

§2. Framing the Existential Import Debate

Aristotle's traditional 'Square of Opposition' is a collection of four general expression forms, and their relations to one another. These general expression forms are traditionally labelled as follows, where S and P are general placeholders for the subject and predicate terms:

A Every S is P
E No S is P
I Some S is P
O Some S is not P (equivalently, Not every S is P)

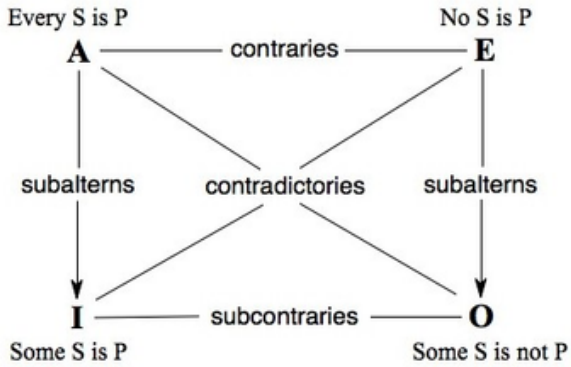
The relations between **A**, **E**, **I**, and **O** are typically represented in a diagram (as shown below). Although these expression forms are often referred to as 'propositional forms', this is misleading as they rather act as schematics for expressions. As such, I will instead refer to these forms as 'expression forms' throughout this essay.

The present debate concerns the existential import of these expression forms. This essay focuses on the two primary approaches to this debate. These

approaches can be characterised as follows:

- **The Classical View:** The affirmative traditional expression forms (**A**, **I**) have existential import and the negative forms (**E**, **O**) do not. This view is typically associated with Aristotle and nominalists after Ockham.[4]
- **The Contemporary View:** The particular traditional expression forms (**I**, **O**) have existential import and the universal forms (**A**, **E**) do not. This is the standard contemporary view.[5]

These are two different interpretations of the four traditional expression forms (**A**, **E**, **I**, **O**), depicted in the traditional square of opposition below:[6]



When examining the practical differences between the two approaches toward existential import, it is pertinent to note that both approaches reach the same conclusion about the form **E** (that it has no existential import) and the form **I** (that it has existential import), and differing conclusions about **A** and **O.**[7] As such, the latter two forms are of particular interest for this essay.[8] I favour Ackrill's translation of the expression form **O**: 'Not every S is P,' rather than the commonly used 'Some S is not P'.[9] In the context of Aristotle's work, Ackrill's translation seems most appropriate because it highlights that **O** is the contradictory of **A**: $Axy = \sim Oxy$ for all x and all y.[10] When understanding **O** simply as **~A**, we see that the extent of the disagreement between the Classical and Contemporary approaches can be characterised as a disagreement regarding expressions of the form **A**.[11] The conclusion one reaches regarding the existential import of A immediately informs her conclusion regarding the existential import of ~A (O). Moreover, the differences only have practical implications in the event that S does not refer; since if S does refer, then $\exists x[Sx]$ & $\forall x[Sx \rightarrow Px]$ and $\forall x[Sx \rightarrow Px]$ have the same truth value. Every differing truth value between the two approaches stems from their different approaches to this particular form. Thus, when we critically compare the two approaches for this specific case (of the form **A** when S does not refer), we effectively compare the

approaches as a whole.

The conclusion one reaches regarding the existential import of A immediately informs her conclusion regarding the existential import of **~A** (**O**). Moreover, the differences only have practical implications in the event that S does not refer; since if S does refer, then $\exists x[Sx] \ \& \ \forall x[Sx \rightarrow Px]$ and $\forall x[Sx \rightarrow Px]$ have the same truth value. Every differing truth value between the two approaches stems from their different approaches to this particular form. Thus, when we critically compare the two approaches for this specific case (of the form **A** when S does not refer), we effectively compare the approaches as a whole.

Considering the especially relevant case—wherein S does not refer to an expression of the form A—both approaches assign particular proposition forms to the same expression form, namely, A. On the Contemporary View (particular-universal approach), one assigns to this expression form the propositional form $\forall x[Sx \rightarrow Px]$, and on the Classical View (affirmative-negative approach), one assigns to this expression form the proposition form $\exists x[Sx] \ \& \ \forall x[Sx \rightarrow Px]$. It is my position that assigning any such proposition to an expression of this general form in OL is done so arbitrarily and ought to be avoided. It is in assigning any general propositional form to an expression of the form 'Every S is P' that both approaches are mistaken.

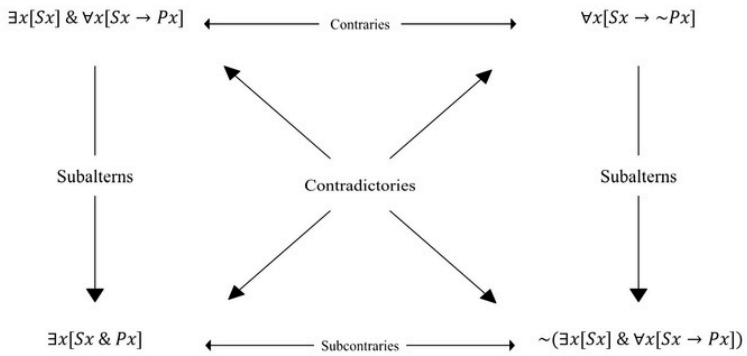
§3. Ambiguity and Nonsense

The existence of the very debate with which this essay is concerned indicates the ambiguity of expressions of the form **A**. 'Every S is P' is ambiguous insofar as two fluent speakers may read it and come to different understandings as to whether or not it implies that there exist things that are S. If we are to properly interpret an ambiguous expression in OL, then we must understand what is meant by the speaker.[12] Consider the following example: Seán owns a pet chicken who has recently eaten her portion of seeds and will not eat again for some time. Simultaneously, Seán is cooking a chicken breast for lunch that has just reached the point where it is safe to eat. We cannot assign a truth value to the expression 'Seán's chicken is ready to eat', without first clarifying what it means; that is, what proposition it corresponds to: Edible(Chicken-Poultry) *or* Hungry(Chicken-Pet). Without disambiguating the meaning of the expression through external context, we cannot decide how we should generally interpret it. In OL, finding out what is meant by an expression is done intuitively, based on the surrounding context. However, in isolated, particular expressions or abstract cases—such as the categorical forms with which we are dealing—there is no such context. Thus, we have no basis for rearticulating expressions of form **A** such that they are propositional signs. In propositional logic,

expressions are useful only insofar as they represent propositions; as such, if an expression is not the sign of some proposition then it is outside of logic's purview. We can of course draw conclusions about the two propositional forms that the two approaches assign to the expression form **A**, but we can only do so if these propositional forms are independent starting points. We simply cannot say anything in logic about the ambiguous expression in and of itself as it corresponds to no particular proposition. That is to say, under our definitions, such ambiguous expressions are nonsense. Thus, any expression of form **A** is nonsense. This highlights the shared mistaken assumption of both approaches to existential import, namely: 'There *can be* a standard method of determining the existential import (directly or indirectly) of expressions of the form **A** that is accurate and superior to other potential methods'.[13] Just as with 'Seán's chicken is ready to eat', 'Every S is P' is ambiguous, and thus nonsense; we cannot make use of such expressions in logic without first finding their intended *meaning*.

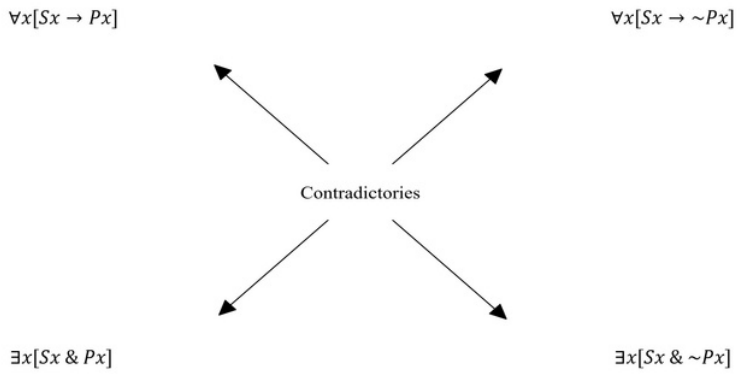
§4. Preserving the Usefulness of a Misguided Debate

Within logic, we may instead draw conclusions about the propositional forms $A^{P1}: \exists x[Sx] \& \forall x[Sx \rightarrow Px]$, and $A^{P2}: \forall x[Sx \rightarrow Px]$. These are distinct propositional forms that are associated with **A**: A^{P1} on the Classical View and A^{P2} on the Contemporary View. Neither A^{P1} or A^{P2} *necessarily* correspond to **A**, and they can be restated in OL as S^1: '*There is some S and if something is S then it is P*' and S^2: '*If something is S then it is P*' respectively. Expressions are 'assigned truth values' only insofar as they correspond to a proposition with a truth value; e.g., 'Purple is equal' has no truth value, and 'Some men are hungry' has a truth value insofar as it represents the truth-apt proposition $\exists x[Mx \& Hx]$. As such, recognising that expressions of the form **A** are nonsense prohibits asking 'Is 'Every S is P' True?' (for any given 'S' and 'P'). Rather, we should seek a clarified propositional sign as a starting point.[14] This leaves us with two coexisting (and indeed useful) 'squares of opposition', as follows:



The Square of Opposition Regarding A^{P1}

In this case we are using the propositional form A^{P1} : $\exists x[Sx]\ \&\forall x[Sx\rightarrow Px]$ rather than **A**. All that Aristotle attributed to the form **A**—its logical relations, laws of inference, and conclusions—in fact applies instead to A^{P1} (insofar as they were well reasoned conclusions to begin with) as this was the meaning with which he was concerned.[15] Unfortunately, these conclusions were tainted by starting from the false assumption that A necessarily means A^{P1}. Therefore, the logical system that lends itself to the Classical View is still useful; we just require the starting point A^{P1}, rather than **A**.



The Square of Opposition Regarding A^{P2}

As in the previous diagram, all the conclusions Russell and contemporary logicians draw about **A** (insofar as they are well reasoned to begin with) remain valid, but instead of pertaining to **A**, they pertain to A^{P2}, and say nothing about expressions of form **A** directly. In the case of A^{P2}, only the contradictory relations hold true.

Conclusion

In conclusion, the debate between which interpretation of **A** is 'correct' or even 'preferable' is misguided; an expression of form **A** is ambiguous. We ought to focus on what conclusions we can draw from propositional signs and their corresponding propositions, and typically try to avoid expressions of form **A**, replacing them with expressions of form S^1 or S^2 according to our intended meaning. Neither way of interpreting **A** is generally favourable. Thus, neither interpretation of **O**—the contradictory of **A**—is favourable. Thus, neither approach to existential import is favourable. Despite this, conclusions that have followed from a given understanding of **A** are still useful for drawing conclusions about propositions of the form that they (mistakenly) assign to **A** in general. Going forward, we should not approach the existential import of the traditional forms in either of the popularly proposed ways, but we should neither discard their teachings, as both are correct, albeit concerned with

different propositional forms or meanings (which they unfortunately both express with 'A').

NOTES

[1] These (perhaps somewhat unorthodox) definitions are required in order to clarify the distinction between ordinary language statements (i.e. expressions) and their meaning (i.e. propositions).

[2] One may further divide this category further into 'ill-formed' expressions and 'ambiguous' expressions (which may represent some proposition *in context*). However, as discussed later in this essay, both sides of the existential import debate are making general—that is, *contextless*—claims. Both the views suppose a superior, general method for interpreting all expressions of the form **A**, **E**, **I** or **O**; albeit for different reasons. Thus, for the purposes of this essay, the ill-formed/ambiguous distinction is not applicable. There is no context with which one may disambiguate ambiguous expressions whilst considering *general* expression forms.

[3] Of course, not all meaningful expressions must be expressible in QL. Rather, I claim only that the expressions relevant to the present discussion can be classified either as 'nonsense' or 'parsable in QL'. Meaningful expressions that are not parsable in QL—such as modal expressions, self-referential expressions, &c.—elude the scope of this essay.

[4] Klima 2008, p.144; Aristotle, 1981, 1011b25; 1989, 13b12

[5] Russell 2009, p.62; Frege, 1879, §12; 1972, pp.134–135. See also (Quine 1948).

[6] Diagram: (Parsons 2021, §1)

[7] They agree on the existential import of **I** because **I** is both positive and particular. **I** meets the criteria of both approaches for existential import. The reverse is true of **E**.

[8] When considering logical structure, the only significant difference between *propositions* is their truth values (though this is not true of expressions). E.g., '(~A) & B' is equivalent to 'A → B' because they share the same truth values.

[9] Aristotle 2020; Parsons 2021, §2.2

[10] (Keynes 2019, p.145). The contradictory relations remain valid on either approach to existential import (see pp.5–6). As such, this translation choice does not favour one approach over another.

[11] Alternatively, the disagreement could be characterised as concerning only **O**.

[12] Here, 'properly interpret in OL' means rearticulating for ourselves the expression such that it is a propositional sign inline with intentions of the expression's originator, rather than nonsense.

[13] Note that the Contemporary View and the Classical View—whilst they both advocate for a superior general method—differ in their reasons for preferring their respective methods. The Contemporary View prioritises convenience and simplicity, whereas the Classical View prioritises the truth conditions of the expression forms. Nonetheless, their shared assumption remains despite their different motivation, and it is this assumption that ultimately fails.

[14] An expression of form **A** is ambiguous even when S refers. However, in this case, AP1 and AP2 would reach the same truth value. Thus, one can clarify **A** in this case, but this would not be particularly impactful. Analogously, 'Michael Collins is dead' could refer to the former astronaut, or the former Irish minister for finance, but it is true either way.

[15] For example, A can be validly obverted to read $\forall x[Sx \rightarrow \sim(\sim P)x]$. (Keynes 2019, p.101)

Bibliography

Aristotle (1938) *Categories. On Interpretation. Prior Analytics*. Translated by H.P. Cooke and H. Tredennick. Loeb Classical Library. London: Loeb.

Aristotle (1981) *Metaphysics*. Edited and translated by W.D. Ross. Oxford: Clarendon Press.

Aristotle (2020) *Categories and, De Interpretatione*. Translated by J.L. Ackrill. Clarendon Aristotle Series. Oxford: Oxford University Press. Available at: https://doi.org/10.1093/actrade/9780198720867.book.1.

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle a. S.: Louis Nebert. Reprinted in Bynum, T.W. (ed. and trans.) (1972) *Conceptual Notation and Related Articles*. Oxford: Clarendon Press.

Keynes, J.N. (2019) *Studies and Exercises in Formal Logic*. Project Gutenberg. Available at: https://www.gutenberg.org/files/59590/59590-h/59590-h.htm (Accessed: 18 February 2026).

Klima, G. (2008) *John Buridan*. Oxford: Oxford University Press.

Parsons, T. (2021) 'The Traditional Square of Opposition', in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition). Available at: https://plato.stanford.edu/archives/fall2021/entries/square/ (Accessed: 18 February 2026).

Quine, W.V.O. (1948) 'On What There Is', *The Review of Metaphysics*, 2(5), pp. 21–38.

Russell, B. (2009) *The Philosophy of Logical Atomism*. Edited by B. Russell. London: Routledge.

On Equivocation in the Frege-Geach Problem

by Patrick Byrne

Expressivism is a metaethical theory claiming that moral language does not state facts about the world but instead expresses an attitude toward its object. For an expressivist, "murder is wrong" means something like "boo, murder!" rather than describing anything in the world. While the reader may find such a view initially unintuitive and objectionable, Byrne masterfully argues that one common objection against expressivism – that it equivocates in the meaning of its terms in *modus ponens* (if X then Y) arguments – is not unique to expressivism at all! While the paper may not convert the reader to expressivism, it powerfully dismantles one obstacle in the way of such a conversion.

---

1. Introduction

In this essay I discuss the Frege-Geach problem for metaethical expressivism, and whether it can be solved by expressivists. First, I investigate one aspect of the Frege-Geach problem: the expressivist treatment of moral modus ponens arguments. A modus ponens argument is one of the form 'P, if P then Q, therefore Q'. It is said[1] that expressivism's account of moral modus ponens arguments is guilty of a fallacy of equivocation, because the same statement is used in two different ways in the course of the argument–'P' in the first premise does not actually mean the same thing as 'P' in the second premise. I first explain all of the above terms in greater detail, then I argue against this idea using a 'companions in guilt' strategy: I argue that the rival theory to expressivism –cognitivism–faces this same problem, and so either both theories or neither should be implicated by it. Then I question what it means to 'solve' the Frege-Geach problem in light of the fact that it targets many more areas of expressivism than just its treatment of the modus ponens arguments that this essay focuses on. I argue that while the problem may never truly be 'solved', it is possible for the problem to be weakened to a point where it is not a devastating objection, and progress in defending expressivism from some instances of the problem–like in modus ponens arguments–gives expressivists good reason to keep working on the problem, even in the face of its current strength.

Metaethical expressivism is the idea that moral statements–like 'Lying is wrong' or 'Keeping promises is obligatory'–are very different from most non-moral statements–like 'Grass is green' or 'Joyce is a talented writer'–in a particular way.

While non-moral statements express beliefs in matters of fact, moral statements have a different purpose: they express approval or disapproval of the thing mentioned in the statement, or perhaps a command to do or not do an act mentioned in the statement. I will use 'expressivism' as a general term for theories which treat moral statements as expressions of desires or commands and I will use 'cognitivism' as a general term for theories which treat moral statements as expressions of beliefs in matters of fact. So, expressivism about moral statements would translate an ostensibly belief-expressing statement like 'lying is wrong' as 'boo lying!' or 'don't lie!'.

2. The Equivocation Argument

The Frege-Geach problem (so-called because it was written about by Peter Geach, who credited the basic idea to Gottlob Frege) criticises expressivism's treatment of moral language, and argues that it fails to explain many common and essential features of moral discourse. Though there are instances of the Frege-Geach problem which target different aspects of expressivist moral language, we will focus on the accusation of equivocation. The equivocation element of the Frege-Geach problem claims that expressivism cannot account for modus ponens arguments with a moral subject matter, while cognitivism can. A modus ponens argument is one of the form 'P, if P then Q, therefore Q'. A typical example of a modus ponens argument with moral subject matter goes as follows:

P1: It is wrong to lie
P2: If it is wrong to lie, then it is wrong to get one's little brother to lie
C1: It is wrong to get one's little brother to lie

On a cognitivist reading, this argument is quite simple. The first premise is taken to be true, the second premise links the truth of the first premise with the truth of another statement, and so the conclusion claims that the other statement is true as well. But the equivocation problem says that complications arise for an expressivist reading of this argument. This is because, for expressivism, the purpose of the first premise is solely to express disapproval of lying, or a command not to lie, but the second premise does not actually express this disapproval or command, it only says what would follow if this disapproval or command was expressed. To put it another way, in the first premise the content of the phrase 'it is wrong to lie' is asserted, while in the second the content of the same phrase is not asserted, it is only discussed. Therefore, an expressivist reading of this argument is guilty of equivocation,

whereby the same phrase is used in two different ways–first asserted and then merely discussed–and so the argument is invalid. Several major metaethicists all directly use the word 'equivocation', making this a serious objection, as if the objection holds, expressism would not simply be unintuitive when it comes to moral modus ponens arguments, it would be fallacious and therefore unacceptable (Blackburn, 1992, p. 190, Miller, 2003, p. 41, Van Roojen, 2024, section 4.1).

In the next section I discuss this objection. Specifically, I ask why this accusation of equivocation should not also be made against the cognitivist reading of this modus ponens argument, as cognitivist readings of this argument also move from asserting the statement in the first premise to hypothetically discussing it in the second. Since cognitivist readings of the argument are taken to be valid, the truth of my objection would necessitate that either the expressivist reading is taken to be valid, or that moral modus ponens arguments are given up altogether.

So, what is the difference between the cognitivist and expressivist readings when it comes to equivocation?

3. Expressivist Content and Companions in Guilt

The most compelling answer to the question of why cognitivist modus ponens arguments avoid equivocation is that cognitivist statements have content that is independent of assertion. There are two distinct roles to a cognitivist assertion: describing a state of affairs, and then asserting that that state of affairs actually obtains in the world. The first premise of our modus ponens argument both describes a state of affairs and asserts that it obtains, while the second conditional premise only describes it. Therefore, the cognitivist modus ponens argument avoids equivocation because, though assertion changes between premises, the description does not.

Alexander Miller, in his *Introduction to Contemporary Metaethics,* uses this strategy. He writes that: "the state of affairs asserted to obtain by [the first premise] is the same as that merely hypothesized to obtain in [the second premise]...Throughout, the semantic function of the sentences concerned is given in terms of the states of affairs asserted to obtain in simple assertoric contexts" (Miller, 2003, p. 42). The function of the sentences is to describe a state of affairs, and assertion can be theoretically separated from that description.

For my companions in guilt strategy to succeed, and for the charge of equivocation to be lifted from expressivism, we must show that an expressivist statement can have some content that is independent of, but can be tied to, assertion. I think this can be done.

Describing this issue, Bob Hale writes that: "I shall say that when a man (sincerely) affirms 'Lying is wrong', he both *presents* a certain attitude (disapproval of lying) and registers commitment to it" (Hale, 1993, p. 338). This idea of 'presentation of an attitude' is what we are looking for. It allows us to hold the idea of disapproving of lying in our heads without committing ourselves to it in the same way that a cognitivist statement allows us to picture a state of affairs without necessarily believing it to be true. (It allows us to discuss the expressivist statement without asserting it). This gives expressivist statements content that is not equivalent to assertion–what Hale calls 'registering commitment'–and it does this while retaining the attitudinal core of expressivism, and therefore not veering too close to the descriptive role of cognitivism.

A counterargument against the idea of the presentation of attitudes might be that the presentation cannot be separated from the registering of commitment to the same attitude–in real speech, nobody can present an attitude without simultaneously committing themselves to it–and so the separation of the presentation from the commitment for theoretical purposes is redundant or illegitimate. I argue, again in line with the companions in guilt strategy, that the same is true of cognitivism: one cannot describe a state of affairs without simultaneously having an opinion about whether it obtains. When a person tells a lie, or utters a statement within a fiction, they know that what they are saying is false even though they are presenting it as true (of course there are differences between lies and fiction, but these differences are not relevant for the purposes of this point). When a person guesses something that they're not certain of, they do not have no opinion about the truth of the statement, they have the opinion that the truth of the statement is–to them–uncertain. So again, if cognitivism is allowed to separate in principle the description from the assertion, expressivism should be allowed to separate in principle the presentation from the commitment.

To return to the modus ponens argument, Hale writes: "Someone affirming [the statement in the conditional] presents the same attitude [as the statement in the first premise], but registers no commitment to it" (Hale, 1993, p. 339). In this

way, an expressivist statement can have the same meaning in both asserted and unasserted contexts by presenting the same attitude, and therefore the expressivist modus ponens argument avoids equivocation, just as the cognitivist modus ponens argument avoids equivocation because the cognitivist statement describes the same state of affairs regardless of assertion.

In line with the companions in guilt strategy, I hope now to have outlined a way in which expressivism can use a parallel strategy to that of cognitivism when it comes to equivocation in modus ponens arguments. Since these options are exhaustive, either they are both acceptable or moral modus ponens arguments as a whole are unacceptable. Since moral modus ponens arguments clearly are acceptable, so must be both cognitivism and expressivism when it comes to equivocation. The success or failure of cognitivism and expressivism relative to one another must therefore depend on other aspects of the theories.

4. Solving the Problem

So far, I have remained neutral on whether we can solve the Frege-Geach problem. I have presented a way to avoid equivocation in expressivist modus ponens arguments, but there are many other elements of the Frege-Geach problem than expressivism's treatment of modus ponens arguments. Mark Schroeder writes that, because of this, "The problem is about as far from being discharged as problems come", since expressivists owe us an account of every single linguistic feature in moral language including "negation, conjunction, disjunction, conditionals, and quantifiers; alethic, epistemic, and deontic modals" and more, and cannot use the tools with which the cognitivist successfully explains them (Schroeder, 2009, p. 265). Is this a reason to doubt that the problem can ever be solved?

Whether we can solve the Frege-Geach problem depends on how strong of an objection anti-expressivists take it to be. If the anti-expressivist demands a satisfying treatment of every issue between expressivism and language, then they will likely never consider the problem solved. But no philosophical theory is perfect. It is likely that even the best theories will have elements that are unsatisfying to some, and must simply be accepted on the basis of the strength of the other aspects of the theory in question–perhaps in the hope that a better explanation will be found later, or that the existing explanation will not seem so unsatisfying once the rest of the theory is accepted. On this model of good philosophical theories, there is hope for expressivism to escape from the shadow of the Frege-Geach problem, even without conclusively solving it in the

way anti-expressivists might want.

If someone's only reason for not being an expressivist is the Frege-Geach problem, they must acknowledge that if the Frege-Geach problem were solved, they would have no reason not to be an expressivist. The proliferation of instances of the Frege-Geach problem in many areas of language other than modus ponens arguments might be a reason to doubt that the problem will ever be solved, but for the anti-expressivist to take seriously this proliferation, they again must acknowledge that if this proliferation was smaller, their objection to expressivism would be weaker. So, success in defending expressivism from one instance of the Frege-Geach problem must go some small way towards refuting the anti-expressivist, or there must be a threshold past which the Frege-Geach problem becomes a less serious doubt (say, credible progress is made by expressivists in 60% of instances). Even if the anti-expressivist believes this threshold will never be reached, it is not irrational for the expressivist to keep trying, and progress in one instance will justify that attempt even further. So, can we solve the Frege-Geach problem? I don't know, but it is not yet irrational for expressivists to keep trying.

5. Conclusion

In this essay, I first introduced the idea of metaethical expressivism and the challenge of equivocation levelled at its treatment of moral modus ponens arguments as a particular instance of the Frege-Geach problem. It is said that these arguments equivocate because in one premise a moral statement is asserted, while in another the same statement is unasserted, and this changes the nature of the statement to a radical extent. I argued that expressivist statements have content that does not depend on assertion by invoking Hale's idea of the 'presentation of an attitude', and that this content is what remains stable between premises. Since the attitude presented remains the same between both premises, there is no equivocation. As an equivalent strategy is used to avoid accusations of equivocation by cognitivists–the expressivist's rival–either both of these theories or neither of them must fail.

Then, I claimed that the demand an anti-expressivist might make for an expressivist to *solve* the Frege-Geach problem may be too extreme, and that with more humble expectations, success in defending expressivism from some instances of the problem give expressivists good reason to keep pursuing their project, even given the current strength of the problem.

I believe there is more to be said about the idea of expressivist content that I have presented here against the equivocation argument, especially given that it does not solve all problems with the expressivist modus ponens argument–for instance, it is unclear if expressivist modus ponens arguments can be called 'true' or even 'valid' given that the first premise and the conclusion are non-cognitivist statements, and therefore not truth-apt.[2] However, I have presented this idea only as an example of a way in which one specific instance of the Frege-Geach problem–the equivocation argument–can be solved, and so a way in which the problem may eventually be overcome.

NOTES

[1] Examples include Simon Blackburn, Alexander Miller, and Mark van Roojen. Cited in more detail below.

[2] This is a problem due to the failure of a "higher-order attitude" account like that in Blackburn's *Spreading the Word*. For the failure, see Schroeder 2008, pp. 708-710.

Bibliography

Blackburn, S. (1992). *Spreading the Word: Groundings in the Philosophy of Language*. Oxford University Press.

Hale, B. (1993). "Can There Be a Logic of Attitudes?" In J. Haldane & C. Wright (Eds.) *Reality, Representation, and Projection*. (pp. 337-358). Oxford University Press.

Miller, A. (2003). *An Introduction to Contemporary Metaethics*. Blackwell Publishing.
Schroeder, M. (2008). "What is the Frege-Geach Problem?" *Philosophy Compass*, Issue 3. URL= https://philarchive.org/rec/SCHWIT-3.

Schroeder, M. "Hybrid Expressivism: Virtues and Vices." *Ethics* 119, no. 2 (2009): 257–309. https://doi.org/10.1086/597019.

Van Roojen, M. "Moral Cognitivism vs. Non-Cognitivism", *The Stanford Encyclopedia of Philosophy* (Summer 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/sum2024/entries/moral-cognitivism/>

Between Right and Wrong: Indeterminacy as a Reply to the Argument from Moral Disagreement

by Jessica Kiehstaller

The everlasting dispute between moral relativists and moral objectivists is approached from a new angle by Jessica Kiehstaller in this essay on moral disagreement. You might hold the position, for instance, that lying to children about Santa is morally permissible. I, however, might wholeheartedly disagree, thinking it is morally reprehensible to deceive them in this way. Neither of us can be persuaded to change our minds. What are we to make of this? Kiehstaller presents some ways we can account for such seemingly irresolvable disagreements, while at the same time retaining objectivist commitments to universal moral truths.

This essay looks at the metaethical debate surrounding moral disagreements and argues for the idea that the indeterminacy of the truth or falsehood of some statements about what is morally right or wrong can enrich this debate. Moral disagreement occurs when two parties give opposing answers to the question of whether a moral statement is true or false. For example, when one person holds that the statement "lying is morally wrong" is true, while another person claims it to be false. My focus here is on the "argument from disagreement," which states that a) a kind of moral disagreement is possible in which neither opposing answer is unjustified (i.e. a rationally irresolvable disagreement), and b) this kind of moral disagreement can only be explained by moral relativist theories (Gowans, 2021, sec. 2, sec. 5). The argument from disagreement is taken to support moral relativism while posing a threat to moral objectivism. I argue instead that a subset of both relativist and objectivist theories can plausibly explain this kind of moral disagreement. Therefore I reject the notion that relativism offers the best explanation simply by being able to provide one.

I will first minimally define moral relativism and objectivism. Then I will illustrate the traditional objectivist explanation of moral disagreements as "rationally resolvable" with a veil of ignorance thought experiment. I will then consider the relativist's challenge in the argument from disagreement that the objectivist's explanation fails to engage with the possibility of rationally irresolvable disagreements. In the rest of the essay, I will explore how objectivists can plausibly explain rationally irresolvable disagreements using the idea of indeterminacy. I will consider three ways in which objective moral facts can fail to be determinate: a) when moral principles are not applied to particular cases, b) two courses of action are morally equal, or c) the moral facts are indeterminate in borderline cases.

Moral objectivism and moral relativism

Moral objectivism is a group of positions which hold that moral statements are true or false objectively, i.e. their truth value holds in all contexts and for all moral agents (Gowans, 2021, sec. 2; Enoch, 2009, p. 16). According to objectivism, an action cannot be morally right and morally wrong at the same time. Similarly, a moral judgement cannot be both true and false at the same time, because there exist objective moral facts that justify only one answer or the other. Therefore, in the case of moral disagreements the objectivist must hold that at most one party's moral judgement is true. This conclusion is the target of the argument from disagreement which I analyse below. Moral relativists, unlike objectivists, accept that the truth of (at least some) moral statements is relative to a context or standpoint, such as an individual or a group (Gowans, 2021, sec. 2). A moral statement is either true or false within one of those contexts, but it may be true or false *only* within that context (and is therefore only true or false relative to it).

The possibility of rationally irresolvable disagreements

The traditional objectivist explanation for moral disagreement claims that at least one party is mistaken in their judgment because of biases, errors in reasoning, or a lack of awareness of relevant facts (Enoch, 2009, p. 25). This means that these disagreements would be rationally resolvable by eliminating these errors or biases. Following the approach of Muldoon *et al.,* (2014) I will use a Rawlsian veil of ignorance thought experiment to illustrate the traditional objectivist response. The veil of ignorance as originally introduced by Rawls is a thought experiment in which individuals choose principles of justice without knowing their position in society, their values or any other facts about themselves. This is meant to ensure that their judgement is completely impartial (Muldoon *et al.*, 2014, pp. 378f.). In the case of moral disagreement, the veil of ignorance eliminates all biases someone judging moral statements as true or false could have and assumes that the 'ideal judge' has no cognitive or epistemic limitations (Muldoon *et al.*, 2014, pp. 378-381). Hence, no moral disagreements would occur behind the veil of ignorance, which shows that they are in principle rationally resolvable disagreements between 'imperfect' judges in real life (Goldman, 2022, p. 851).

Relativism, on the contrary, argues that some disagreements appear (at least intuitively) to originate from something more than biases or errors; it is at least *imaginable* that there are some disagreements which would not dissolve behind the veil of ignorance (Muldoon *et al.,* 2014, pp. 381f.; Tersman, 2022, sec. 1.3). Two courses of action might each seem morally right (from a respective

standpoint) and choosing one over the other as the universally "morally right" one would seem arbitrary (assuming there is no independent standard that the ideal judges can appeal to). Relativism can explain this (rationally irresolvable) disagreement by saying that the ideal judge's basis for choosing – i.e. their context and social background – has been eliminated. Since the veil of ignorance thought experiment is set up such that, for example, the society one belongs to is not morally relevant, the ideal judge is not aware of any 'belonging'and is not able to choose one action as the right one. For the relativist, a moral judgment being true or false *relative to a context* is the best explanation for some disagreements (potentially) not being rationally resolvable.

The argument from disagreement

The argument from disagreement challenges whether *all* moral disagreements can be explained 'away' as rationally resolvable (Enoch, 2009, p. 24). It seems that moral facts do not give us one universally holding answer to instances of this kind of disagreement. This issue arises for objectivists in particular arguably because they adopt a stricter principle of "bivalence" (Shafer-Landau, 1995, p. 83) than relativists. Both strands of theory (relativist and objectivist) deal with morality as needing to be able to provide a determinate answer for whether a moral statement is true or false. But the relativist can deal with instances of rationally irresolvable disagreement because they can respond that there can be only one determinate answer *within* or *for* one society, moral community (or otherwise defined context or standpoint). However, for the objectivist, this principle must hold universally, i.e. if a moral statement is true, it is true in every context and for every group.

The argument from disagreement is conventionally presented as an inference to the best explanation, as is the case in Alan H. Goldman's (2022) formulation. The argument takes the possibility of rationally irresolvable disagreements seriously, given the deep disagreements that seem rationally irresolvable in real life (Enoch, 2009, p. 39). The *hypothesis* that ideal judges would come to a conclusion in cases about which there is deep disagreement in real life is arguably not enough to rule out the possibility of (some) disagreements being instead rationally irresolvable (Enoch, 2009, pp. 39f.; Goldman, 2022, p. 857). To engage with this challenge and this possibility seriously, moral objectivists need to either a) explain how rationally irresolvable moral disagreements are compatible with there being objective moral facts (about those matters), or b) deny that there are any objective moral facts (relativism) (Goldman, 2022, pp. 850-853). Goldman holds that for some disagreements explanation b) is the only one available and is therefore the best one (p. 853). Three inconsistent

claims seem to be in play in the argument from disagreement:

1. Moral statements are objectively and universally true or false.
2. Some rationally irresolvable disagreements are imaginable.
3. All moral judgements must be determinate within a moral system (whether a relative or universal one).

Relativists accept '2' and they can accept '3' as well, since they can interpret a "moral system" in '3' as not universal, given that they deny '1'. Objectivists typically accept '1' and '3' but deny '2'. However, objectivists seem to be forced to at least consider '2' in order to face the challenge of the argument from disagreement. But to make their position consistent, they would need to deny '3', since they cannot give up '1' (which is a cornerstone of objectivism). In the rest of the essay, I will explore two ways in which the objectivist can deny '3' in some cases (without giving up '1') and thereby account for '2' (the possibility of rationally irresolvable disagreement).

Underdetermined moral principles

The first way in which a moral judgement might not be determinate takes place at the level of moral principles. In the example of judging different principles of distributive justice, the question arises whether the principle of merit-based distribution or distributing based on need leads to a morally right action (Muldoon *et al.,* 2014, p. 378). Assuming that neither of the principles is *objectively and universally* 'better,' it seems that an ideal judge could not choose either principle as the correct one without their choice being in some way arbitrary. However, ideal judges *could* come to a determinate answer on whether distributing a good 'G' in a *particular situation* 'S' based on merit is morally right while distributing 'G' not based on merit in 'S' is wrong. Considering the particular constituents of the situation 'S' might determine how the principles need to be weighed, or which principle applies in situation 'S'. So, while there is no determinate answer on the level of principles in general, a determinate answer can be generated by ideal judges regarding the status of rightness or wrongness of an action in a particular situation involving these principles (Brink, 2000, p. 159f.). The same moral statement – "distributing 'G' based on merit is morally right" – can be true in a situation 'S' and false in a situation 'S*'. But just looking at the statement "distributing 'G' based on merit is morally right" seems to be neither determinately true nor determinately false in and of itself. Nevertheless, the statement has a determinate truth value when applied to situation 'S' or 'S*' for example. And in situation 'S' the statement is not only determinately true, but it is true objectively. The truth or falsehood of

the statement is not relative to a group or context, but only relative to the situation in which the principle is applied.

One possible (relativist) challenge to this is that the question of how to weigh principles in particular situations is not a matter of impartial rationality alone. They might be additionally weighed according to some (disputable) standard (Muldoon *et al.*, 2014, pp. 381ff.; Goldman, 2022, pp. 857). But this objection is not convincing in relation to the considerations at hand, since it might suggest that there needs to be not only meta-principles to decide how to weigh principles (such as Utilitarianism) but also meta-meta-principles for those and so on. It seems more plausible that under the objectivist assumption of objective moral truths, there is only one correct weighing of principles in a particular situation, which *ideal* judges would correctly determine (Goldman, 2022, pp. 861f.). Whether there is one correct meta-principle for weighing them and which meta-principle that would be is irrelevant here. The relevant conclusion for the objectivist position is that there is an objectively correct weighing given the particular constituents of a specific situation and that ideal judges would be able to determine and apply it in their judgement.

Indeterminate moral judgments

To extend the scope of rational disagreements which can be successfully explained by objectivists, I will explore two ways in which the application of moral principles in a particular moral judgment could (in some cases) fail to generate one determinate answer about what is morally right or wrong.

a) Moral conflicts in which all facts are equal

Moral indeterminacy could arise in a situation where all moral and relevant non-moral facts are the same for multiple potential actions. There would then be more than one morally permissible action (Goldman, 2022, pp. 860). For example, if you can only save one of two people 'C' and 'D' from a burning house, there is no determinate answer as to which one of them it is morally right to choose to save, provided there is no morally relevant difference between them (Goldman, 2022, p. 860). In this example "saving 'C' and not 'D'", and "saving 'D' and not 'C'" would both be morally right or at least permissible, even though they are conflicting. So, in this case it is not possible for the ideal judge to decide which course of action is morally 'righter' and should be carried out. Choosing either course of action would be objectively right or permissible, and not because each of the choices of action is morally right relative to different contexts. This kind of indeterminacy is 'weak'. A disagreement about a situation like this one would not be rationally resolvable

by determining one course of action to be the only morally right one, but it would be resolvable by determining that both options are morally permissible. This kind of indeterminacy, as well as the conclusion that both actions are morally permissible, could be accepted from both an objectivist and a relativist perspective. More importantly, this case illustrates that an objectivist position can easily accommodate this 'weak' kind of indeterminacy, which weakens the objectivist position's ties to the determinacy-requirement of claim '3' from above.

b) Indeterminate thresholds

An alternative strand of explanation for the compatibility of moral objectivism with a lack of a determinate answer to some moral statements' truth-value is to take some degree of indeterminacy to be part of some (objective) moral facts' truth values themselves (assuming there are objective moral facts) (Shafer-Landau, 1994, p. 333). This third set of indeterminate cases are ones in which two opposing actions are not equal but both of them seem potentially morally right or wrong at the same time. For example, moral statements about "being significantly harmful": A person 'E"s right to do some action 'A' ends when it *significantly harms* another person 'F'. The "when" implies that there is some *threshold* that must be reached or passed for some action 'A' to count as "significantly harmful". Usually, it is clear if the threshold has been passed or not been passed (Shafer-Landau, 1995, p. 86). But in some "borderline cases" (Shafer-Landau, 1995, p. 88) things become less clear. In a borderline case, one could neither be certain that 'A' causes enough harm to limit 'E"s right to do 'A' nor that it doesn't, no matter how closely the level of 'harm' is analysed in the situation. The judges behind the veil of ignorance would not be able to come up with a determinate answer in a borderline case, because some moral facts might include some slight indeterminacy around this postulated threshold (Shafer-Landau, 1995, pp. 88f.). This kind of 'disagreement' between the imperfect as well as ideal judges would be rationally irresolvable, since there is no moral basis for favouring either answer. Ideal judges might, however, converge on the judgment that the answer is indeterminate (Shafer-Landau, 1994, p. 343).

The potential for moral facts to be indeterminate behind the veil of ignorance is open to doubt, because it seems difficult to imagine that ideal judges could not give a determinate answer. In the same vein as the judges being able to impartially weigh different principles to find a determinate answer, it should in principle be possible for them to know the exact threshold past which something 'becomes' impermissible. It seems intuitive to assume that

ideal judges would have such insights if there are moral facts. However, the way I introduced this kind of moral indeterminacy above is that the indeterminacy is part of the moral facts, not a lack of knowledge of the judges (Shafer-Landau, 1995, p. 84). So, if some objective moral facts do not include information about the exact point at which the threshold for "significantly harmful" is reached in a borderline case, an ideal judge could not know that threshold either. The threshold, rather than being clear-cut, might be a broader area, so that situations falling into that area cannot be clearly labelled as either passing or not passing the threshold (Shafer-Landau, 1995, pp. 84f.). Moreover, in those cases of moral indeterminacy, defining a clear-cut threshold for the sake of having one would not be justified by the moral facts.

A relativist might deal with the case of an unclear threshold by arguing that one person/group could hold that the threshold is passed, and another person/group could hold that it is not passed, since there is no objective fact that states which option is ultimately true (Goldman, 2022, p. 861). If we accept that there can be genuinely indeterminate cases, the relativist's explanation loses the nuance that neither of the statements can be determinately true or false universally (Schiffer, 2002, p. 303). A relativist's explanation would make a morally indeterminate case determinate relative to some context. While this might be a guide for what to do in practice in some indeterminate cases, the moral situation should still be accepted as morally indeterminate between some options. This would also explain *why* it seems possible that a moral judgment could have different objectively justified interpretations independent of any context or standpoint, i.e. because it is actually indeterminate (Wong, 2023, p. 27ff.). Therefore, one reason to favour an objectivist approach to indeterminate cases is that it includes the acknowledgment that in such a case neither moral judgement is nor can be morally true.

Conclusion

In the previous two sections I argued that the objectivist can add a third option c) to the list of possible explanations of moral disagreements (Goldman, 2022, p. 864). With this third option added, moral disagreements are either:

a) rationally resolvable
b) rationally irresolvable because there are no objective moral facts
c) rationally irresolvable because some objective moral facts are indeterminate in borderline cases

Since in the above list there are two options which can explain rationally

irresolvable moral disagreement, the relativist's explanation b) is no longer automatically the best one if we assume that rationally irresolvable disagreement is possible. Three different ways in which a moral judgment could be indeterminate have been discussed and deemed plausible ways to explain moral disagreement from an objectivist view. They cover a range of different possible cases, although it has not been proven that they cover all potential cases of rationally irresolvable disagreement. What has been established is only that if a disagreement is genuinely irresolvable *and* there are objective moral facts *and* these moral facts can be indeterminate, then this disagreement may be irresolvable because it is a borderline case and the moral facts in question are indeterminate for that borderline case. At the end of the previous section, I briefly sketched one reason to favour the objectivist explanation of borderline cases as more nuanced and inclusive. However, the explanation which is ultimately the correct one depends on whether one accepts that there actually are (indeterminate objective) moral facts. Nevertheless, after adding c), the objectivist explanation cannot be dismissed as incomplete and needs to be at least considered along with the relativist's explanation b) for the possibility of irresolvable disagreement.

Bibliography

Brink, D. O. (2000) 'Moral disagreement', in C. W. Gowans (ed.) *Moral Disagreements: Classic and Contemporary Readings*. Routledge, pp. 157-167.

Enoch, D. (2009) 'How Is Moral Disagreement a Problem for Realism?', *The Journal of Ethics*, 13(1), pp. 15–50.

Goldman, A. H. (2022) 'The Right Argument from Moral Disagreement', *Theoria*, 88(4), pp. 850–867.

Gowans, C. W. (2021) 'Moral Relativism', in E. N. Zalta and U. Nodelman (eds.) *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition). Available at: https://plato.stanford.edu/archives/spr2021/entries/moral-relativism/ (Accessed: 4 October 2025).

Muldoon, R. et al. (2014) 'Disagreement behind the veil of ignorance', *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 170(3), pp. 377–394.

Schiffer, S. (2002) 'Moral Realism and Indeterminacy', *Philosophical Issues*, 12, pp. 286–304.

Shafer-Landau, R. (1994) 'Ethical Disagreement, Ethical Objectivism and Moral Indeterminacy', *Philosophy and Phenomenological Research*, 54(2), pp. 331–344.
— (1995) 'Vagueness, Borderline Cases and Moral Realism', *American Philosophical Quarterly*, 32(1), pp. 83–96.

Tersman, F. (2022) 'Moral Disagreement', in E. N. Zalta and U. Nodelman (eds.) *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition). Available at: https://plato.stanford.edu/archives/fall2022/entries/disagreement-moral/ (Accessed: 12 October 2025).

Wong, D. B. (2023) *Moral Relativism and Pluralism*. Cambridge University Press. Cambridge Elements/Ethics.

Should We Be Polite?

by Gavin Dunphy

Mind your P's and Q's! In this essay Gavin Dunphy investigates the relationship between politeness and ethics. We begin by considering an argument put forward by Sarah Buss in which being polite is declared a moral duty. Then, Dunphy respectfully pokes a hole in Buss's argument, showing it to be circular, and expounds a more nuanced argument for politeness. This essay makes a compelling case that while the rules of politeness themselves have no moral significance, in many situations we do have a broader ethical obligation to be polite.

Introduction

In this essay I examine three different positions on the relationship between ethics and codes of politeness.[1] I first describe a position that identifies politeness with morality (section 1), claiming that we have a moral duty to be polite as this is how we show "respect" for others as ends. I will reject this option, showing a circularity in its conception of politeness. I will then outline the response that politeness should be wholly rejected in favour of ethics, motivating it with three objections to the first view (section 2). I finally argue that this second view is unsatisfactory, by giving an argument that while politeness has no ethical force by itself (it is amoral), it is a tool which we very often have an ethical obligation to use (section 3).

It is worth mentioning the origin of this essay, which helps to explain the reason for its structure. It was written for an Ethics class with Damien Storey of Koç University, Istanbul, in response to the paper by Sarah Buss cited in the next section, which he put on the course to discuss why he thought it was wrong. The position discussed in section 2 is a caricature of his objections to the conflation of politeness with ethics articulated to me in various conversations, which he felt were underdiscussed in the literature (explaining my lack of citations). I agreed with the rejection of the sentiment represented by Buss but thought that an argument could be made against a complete rejection of politeness, hence my position in section 3, with which I think he would broadly agree.

1. Arguing for politeness

Sarah Buss (1999) argues that "the 'respect' and 'dignity' of such importance in moral philosophy are the very same 'respect' and 'dignity' of such importance in manners", that "playing [an essential role in our moral life] is the essential function of good manners" (1999, p. 795). We have a moral duty to respect other people, and manners are the part of ethics that deal with many of the important parts of enacting respect. She is giving a philosophical explanation for our "not [being] committed to a clear division of labor between rules of manners and rules of morality" (1999, p. 796). I will demonstrate what I think is a major flaw in this argument, before looking at two ways of moving past it. I stress that while Buss makes an explicitly Kantian case for her position, it is not deontology that I object to, but the contention that the "essential function of good manners" is a moral one.[2]

2. Rejecting politeness

An obvious objection to this view of politeness is that many codes of politeness encourage unethical behaviour by institutionalising disrespect of certain people. There are many uncontentious examples of this, the Indian caste system being a particularly salient one. Buss accommodates this objection by proposing that codes of manners that prescribe disrespect are codes of "bad manners" (Buss, 1999, pp. 810–11). Developing her view she provides an example in which there seems to be a tension between manners and ethics: the custom of men opening doors for women, a sexist form of politeness which may or may not be considered good manners. We can either (1) understand this as good manners which are unethical for reasons outside the domain of politeness, or we can (2) understand this as really being a form of disrespect and thus bad manners (1999, pp. 811–12). My objection is that both options seem to be begging the question—why must manners be respectful of people as ends? Manners are culturally relative and variable enough that it is plausible to suppose that they are as neutral as any other cultural phenomenon with no explicit relation to an objective ethics, such as cuisine. If we are objectivists, and care about respect, say, as an ethical principle, then all cultures *should* recognise the importance of this notion, but it is quite clear that respect can be shown in very different ways in different cultures. It is not given that the 'good' of manners should simply be identified with the 'good' of ethics. If we refrain from this assumption, Buss' argument is shown to be circular: good manners are those through which the ethical imperative of showing respect for others is

fulfilled, and codes where this respect is shown are good manners. Codes of manners that don't meet the ethical requirements are opposed either from an *ethical* position further abstracted from behaviour than the code of manners (1), or from the point of view of good manners, which have already been defined as equivalent to *ethical* behaviour (2). It's ethics all the way down.

Buss *defines* (good) manners as equivalent to an aspect of respecting people, that is, as equivalent to an aspect of her conception of ethics. She uses her ethical commitments to distinguish between different codes that, without an outside source of value (ethics) would not be distinguished, giving an implausibly circular account of the function of manners. This leads to one potential response: there is never an ethical imperative to act politely, just an imperative to act ethically. Imperatives of politeness are often distinguished from those of morals (McPherson, 2019, citing as an example Foot, 2002, pp. 160-166). The 'should' in the title of this essay is a moral one. And though we *should* do what moral imperatives require of us (this is a tautology), while this "should" may sometimes overlap with the "should" of etiquette, no moral significance ought to be given to the latter. We should (morally) do what we should (morally) do, to reiterate the same tautology, but we needn't assent to the claim that we should (morally) do what we should (politely) do. If and where the two shoulds overlap, the moral one is already doing all of the work, and where they contradict the moral one should take precedence because it is moral, and not because of Buss' implausible idea of relabelling the contradictory "should" of politeness as 'bad manners'. In line with this response that moral imperatives supercede imperatives of politeness, we can formalise three objections to Buss' view:

> (a) Politeness is highly culturally relative and seems to be defined contingently within each culture, unlike ethics. This is a seemingly robust empirical claim which I will not attempt to defend (see Culpeper, 2011).
> (b) Imperatives of politeness sometimes necessitate insincere, pointless, or potentially harmful acts.
> (c) Politeness can obstruct moral behaviour, as when it requires us to speak respectfully when we should instead be expressing outrage, or when it encourages unethical behaviour, as discussed above.

While these objections may have weight, I think that our response should not be that politeness ought to be rejected in favour of morality. I will argue for a third position against both Buss and the aforementioned response which rejects politeness based on the above objections.

3. **A more nuanced argument for politeness**

I propose that whatever politeness is, it is a more or less ethically neutral cultural phenomenon (contra Buss), but that there is nonetheless a (qualified) ethical imperative to be polite. We should be polite in ordinary social circumstances, that is, when there is no more pressing ethical imperative. The main argument for this is that

(1) There are morally significant notions relating to ordinary communication, such as "social pain" (Brownlee, 2024, pp. 701–02).
(2) We have ethical obligations under ordinary circumstances to fulfil certain duties with respect to some of these notions, an area that can be called "interactional ethics" (Brownlee, 2024).
(3) Many of these obligations cannot be fulfilled without using subtle, non-propositional forms of communication such as politeness—in the majority of circumstances, unless we are polite, we will fail to fulfil our ethical obligations.

The argument for (3) being that the requirements of politeness are understood throughout any given culture, thus its use and disuse is an unavoidable necessity in communicating in such a way as to fulfil our obligations in (2). Here, I demonstrate that this proposal can respond successfully to the objections outlined above.

Consider a case where someone who is otherwise fulfilling most of their ethical obligations is often rude.[3] Rudeness in very many cases causes offense. This seems to be modified but not extinguished by increased familiarity with the rude person—it is quite conceivable that someone intimately familiar with them persists in taking offense, that is to say, the rude person persists in failing to adjust their behaviour so as not to cause social pain. We can say that this is unethical simply because causing pain without justification is unethical, or because the corrosiveness which goes along with persistently causing pain in a relationship is inimical to the good life (which requires healthy relationships). Either way, we say it is unethical and give ethical arguments for this judgment. The way that this person should change their behaviour is from 'unethical' to 'ethical', but this is an impoverished description –a better one is that they should be polite. They violated an ethical imperative by not being polite, so the ethical imperative that they should have followed is an imperative (in this, ordinary circumstance) to be polite. There is also a pragmatic element here: there are many means by which politeness is taught and practiced in a culture, making the "be polite" imperative more concrete and realisable than an abstract

"be ethical".

With this formulation the response to objection (a) becomes clear. The fact that politeness is hugely or even entirely contingent is of no concern, as it nevertheless has significant force over non-contingent ethical notions such as social pain. Any given instantiation of politeness may be ungrounded, but a failure to follow its imperatives, in many circumstances, is equivalent to failing to follow ethical imperatives. You cannot *just be ethical*, because being ethical necessitates engaging to a significant extent in contingent cultural practices not directly imbued with ethical content.

Objection (b) concerns pointless or insincere behaviour, for example when someone is "just being polite" such as in pretending to be interested in what someone else is saying. The objection could be developed further to argue that if we are genuinely interested in what someone is saying we will appear interested, and if we are not interested it is dishonest (and thus unethical) to use social tools to appear interested. I will push back on both points. If I am genuinely interested in something I am hearing on the radio it is perfectly permissible for me to check my phone, which I am also interested in. If I am equally interested in what is said to me by a person, it is unethical to check my phone because it is rude to check my phone, even if that would be my most natural behaviour. It is rude because it is a failure to communicate my interest in a mutually understandable way and thus risks causing social pain. If I am sincerely uninterested, an enthusiastic expression of interest would be insincere, and potentially unethical, but failing to act politely would also appear to be unethical. Respecting the imperatives of politeness by maintaining eye contact etc., while happily taking any opportunity to detach oneself from the social interaction, seems the option which is most ethical *because* it is polite.

Objection (c) is what motivates the distinction Buss makes between good and bad manners. We will avoid that distinction: politeness for us is entirely contingently defined within a given culture and is not an ethical notion, so the politeness of the caste system really is politeness, though it is quite clear that there is an ethical obligation to violate this code of politeness. We now require a clearer understanding of what I mean by an imperative to be polite in 'ordinary circumstances'. To give an abstract definition, ordinary circumstances are those in which acting politely directly or indirectly fulfils, or leads to the fulfilment of, ethical imperatives.[4] We *should* act to rid ourselves of unethical versions, or elements of, politeness (such as the caste system), and we *should* suspend politeness when it gets in the way of clearer and more important ethical imperatives that occur to us (such as in arguing forcefully with a bigot, or violating someone's space out of medical necessity), and we *should not* make

life miserable for our loved ones, or unpleasant for strangers, by being consistently impolite. The more complicated cases I leave to be worked out by those indefatigable normative ethicists.

I will briefly mention two positive arguments for the view of the role of politeness in our ethical obligations that I have given. One is that in many cases politeness is a prerequisite for meaningful social interactions, *and that* many of these meaningful interactions are required for the good life. This is a variation on the theme of what has been said above, and is briefly discussed by Buss with reference to Hume (Buss, 1999, pp. 798-99). If there is some kind of obligation to engage in certain kinds of social interactions, and being polite is a necessary condition for any of these, then there is a qualified ethical imperative to be polite.

The other point is the role of manners in moral education. There is a strong sense in which the ritualised practice of something (politeness) is very effective in changing our attitudes and thus is highly effective in teaching something, especially something as subtle as morals. Buss points out the comparison of the manners-moral education relationship with the (very effective) use of ritual in religious practice (1999, p. 795), and Berninger (2021) makes a good case that the means by which rituals serve this kind of function is in how they get us to direct our attention. Habitually saying 'thank you', for instance, makes us habitually remember what other people are doing for us, and perhaps even makes us habitually grateful. An objector could of course say that either manners don't teach us in this way, or they can teach us harmfully (as evidenced by how well people learn what Buss calls bad manners). I would reply that this is because politeness is not ethical, it's a social tool.

4. C**onclusion**

I have argued that there is no general ethical imperative to be polite, that politeness is not a type of ethics, and that it is wrong to call something good or bad politeness except by its own contingent, amoral rules. I have also argued that a reaction to this by the removal of any ethical imperative to be polite, as opposed to just being ethical, is wrong. While there are serious problems with the notion that politeness is simply a region of ethics, and that being polite is good for its own sake, I have made a limited positive argument in favour of some sort of ethical obligation to be polite. In circumstances where there is no more pressing ethical imperative, we have an obligation to treat friends and strangers in such a way as to avoid causing them pain. To do this, it seems, we have to engage with politeness. This gives us an obligation to be polite, without

any commitment to politeness as inherently ethical.

NOTES

[1] Taking both concepts broadly but not imprecisely. Ethics is what relates to how we ought to act (remaining agnostic about competing normative ethical positions, and about metaethical positions except for that of objectivism, a presupposition of the kind of discussion I am engaged in here). Politeness, here interchangeable with manners or etiquette, is the highly variable, but almost always present, set of norms, rules, or expectations in a culture concerning the arbitrary aspects of interactions between people.

[2] While a consequentialist account seems less likely (for whatever reason) to take this kind of position, there is nothing to stop the same conflation being articulated with respect to consequences rather than the necessity of respect, as the locus of moral import.

[3] There is a distinction made in some of the literature between politeness and impoliteness (see Culpeper and Haugh, 2021), it is not relevant here because rudeness can refer to impoliteness, but it can also simply refer to a failure to be adequately polite.

[4]I avoid having to make this definition less abstract because in this essay I am not committing to any ethical position aside from objectivism.

Bibliography

Berninger, A. (2021) 'Manners as Desire Management', *The Journal of Value Inquiry*, 55(1), pp. 155–173.

Brownlee, K. (2024) 'On the Ethics of Interacting', *Journal of Applied Philosophy*, 41(4), pp. 697–713.

Buss, S. (1999) 'Appearing Respectful: The Moral Significance of Manners', *Ethics*, 109(4), pp. 795–826.

Calhoun, C. (2000) 'The Virtue of Civility', *Philosophy & Public Affairs*, 29(3), pp. 251–275.

Culpeper, J. and Haugh, M. (2021) '(Im)politeness and Sociopragmatics', in M. Haugh, D.Z.

Kádár, and M. Terkourafi (eds) *The Cambridge Handbook of Sociopragmatics*. 1st edn. Cambridge University Press, pp. 315–339.

Culpeper, J. (2011) '13. Politeness and impoliteness', in G. Andersen and K. Aijmer (eds) *Pragmatics of Society*. De Gruyter, pp. 393–438

Foot, P. (2002) *Virtues and vices and other essays in moral philosophy*. Oxford : New York: Clarendon Press; Oxford University Press.

McPherson, D. (2019) 'Manners and the Moral Life', in T. Harrison and D.I. Walker (eds) *The theory and practice of virtue education*. London New York: Routledge, Taylor & Francis Group.

Two Qualitatively Different Kinds of Pleasure: Imagination, Disinterest, and Free Play in Kant's Aesthetics

by Miriam Treitinger

Are all pleasures the same to us? Can the pleasure, say, of discovering the vacancy of your favourite desk in the Ussher be comparable––*in any way*––to, say, the pleasure of stumbling into a self-sustaining contemplation of Turner's watercolours in the National Gallery? It is *essential* for Kant's aesthetics that the judgments involved in each experience are distinct: as interested and disinterested, respectively. Miriam Treitinger helps Kant by suggesting that an essential difference between these judgements is also the *kind of pleasure* involved. She argues that there is a *qualitative* difference between the 'Ussher' pleasure and the 'Turner' pleasure. And for all this, we need that blind and mysterious function of the soul, which is nothing but the *imagination*...

Introduction

Terminology taken from Immanuel Kant, such as 'disinterested interest', is frequently used in discussions about aesthetics specific to the experience of nature (Budd, 2005, p. 118). In Kant's work, however, disinterested/interested attitudes do not strictly map onto contemplation of the non-human/human world. This essay suggests that one way in which these attitudes differ instead is in terms of the kind of pleasure they are associated with. Working within Kant's system of critical philosophy, I argue that it is the faculty of imagination in particular that, in its harmonious 'free play' with the understanding, allows for a qualitative difference between the pleasure occasioned in judgements of agreeableness (which are associated with interest) and that related to judgements of taste (which are associated with disinterest). After demonstrating that there is such a qualitative difference, I discuss the imagination's distinctive role in bringing about what I call 'disinterested pleasure'.

1. Nature and 'Free Beauty'

Kant distinguishes between 'free' and 'adherent' beauty (Kant, 2000 [1790], p. 5:229). The former is the kind of beauty we attribute to an object when we are either not aware of what kind of thing it is supposed to be, or if we abstract the judgement from our knowledge of the concept (pp. 5: 229–230). Beauty is free when it is the result of a *pure* aesthetic judgement that is entirely 'disinterested' (i.e., a judgement which is not related to anything an object could do for you, but is instead a response to merely the form of the object) (pp. 5: 204–5: 206, 5: 231). As long as we do not have an idea of what a flower is supposed to be, any judgement of it as beautiful is necessarily free (p. 5: 229).

But since we might have an idea of what a human should be like, i.e. a standard that we might refer to, a judgement of the beauty of a person is 'adherent' if we do not disregard our ideas about the concept 'human' in general in the process of making the judgement (p. 5: 230). This kind of judgement is, at least partially, interested (pp. 5: 230–5: 231).

In contrast to the aesthetic appreciation of human-related things, which is typically adherent and hence impure, the aesthetic appreciation of (non-human) nature is typically pure. But the conclusion that free beauty is intrinsically connected to nature does not follow. We may judge anything that we do not evaluate against a standard as beautiful in this way. Having clarified that pure aesthetic judgements are not necessarily about nature, and thereby pre-empted a possible misunderstanding of the subject matter, we can now begin our investigation of the nature of these kinds of judgement.

2. Two Qualitatively Different Kinds of Pleasure

According to Paul Guyer's reading of the *Critique of the Power of Judgement*, pure aesthetic judgements generally rely on "reflection on the context and cause of the feeling of pleasure […], or on its particular 'relation to representation'" (Guyer, 1997, p. 103). These judgements would then be notoriously hard to make, or at least to properly identify within oneself. They seemingly require a subject that is completely aware of the presence or absence of any personal desires or interests, and that is able to determine whether any pleasure experienced in the contemplation of an object is related to these. We would have to accept this to be the case if we, like Guyer, think that Kant does not believe there to be a qualitative difference between the pleasure brought forth in relation to interest and that which results from pure aesthetic, disinterested, contemplation (Guyer, 1997, p. 103). I am rejecting this view. It follows from this rejection that there is an additional, non-reflective, way of determining whether something is beautiful – namely by attending to the quality of the pleasure one experiences in the contemplation of an object.

The difference between the mental acts concerning 'interested' and 'disinterested' judgements seems to necessarily entail a difference in the experience of pleasure. When pleasure is combined with interest, it either pleases the senses, or through reason (p. 5: 206–5: 209). What combines both seems to me to be the presence of a desire that is either instantiated by the object of contemplation itself or is pre-existing. Whether the experience of an object results in the gratification of the desire and therefore depends on the state of the subject (whereby the state itself may be caused/influenced by the object of contemplation). The intensity of pleasure is therefore dependent on

the intensity of the desire. A meal, for example, may taste better the hungrier one is. This kind of pleasure cannot be stable, since prolonged exposure to the object goes hand in hand with a decrease of the pleasure it can provide.

The second kind of pleasure, that which is connected with 'disinterested' judgement, arises from a reflective process that (I argue below) is connected to a particular interaction of the imagination and the understanding. Unlike the first kind of pleasure, it is completely distinct from cognition (which, for Kant, is the act of applying concepts to an object and thereby gaining awareness of the object as such) (pp. 5: 203–5: 204). In other words: we do not need to know the object in order to judge that it is beautiful.[1] Any interest that has to do with the proper judging of an object as beautiful originates from the contemplation of the object's mere form. As long as we assume that the mental faculties of all rational beings work in the same way (something that is intrinsic to Kant's system), any object that we can properly judge as beautiful will elicit a pleasurable reaction in any rational being observing it with the appropriate attention, and the same rational being will experience the pleasure at any time it observes the object with the appropriate attention (pp. 5: 213–5: 216). As the only precondition to this kind of experience is our mental set-up (that we have by virtue of being rational agents), it necessarily follows that, even after long times of observation, as long as proper attention is being paid, the same subject will experience the same pleasure in the contemplation.

A conceivable objection to this distinction is that it might essentially be the same feeling which is experienced in both cases at different intensities and for different amounts of time: the difference analysed here might perhaps be quantitative rather than qualitative. My reply to this objection would be that it makes the most sense to refer to the difference as qualitative, insofar as the difference is apparent from immediate experience. This applies even if the experience differs merely in the intensity of the same feeling or the change of the intensity through time. If the experience of pleasure can be divided into (at least) two distinct experiences that can be distinguished without taking into consideration anything besides the experience, I believe it makes sense to speak of two distinct kinds of pleasure.

Kant's statement that "however dissimilar the conceptions of the object, be they proper to the understanding or even to the reason instead of the senses, the feeling of pleasure […] is always the same" may be interpreted similarly (Kant, 1956 [1788], p. 23, cited in Guyer, 1997, p. 104). Not only does Guyer point out that "the passage was written before Kant finally adopted the view that the feeling of pleasure could be occasioned by the faculty of judgement [and] by sensuous or rational desire" (Guyer, 1997, p. 104); but this may, again,

be interpreted as asserting that all pleasure is essentially the same feeling, while not simultaneously ruling out the possibility that it may be experienced in at least two distinct ways, according to what brings it about.

3. The Role of the Imagination in Bringing About Disinterested Pleasure

So far, I have argued that there are two different kinds of pleasure. I have derived this primarily from the difference between interested and disinterested judgements, but I have only briefly touched upon the fact that these different qualities of pleasure are an immediate result of the respective roles that the faculties of the imagination and the understanding have in creating them. The first kind of pleasure (the 'interested' kind) relies primarily on the understanding, which allows us to turn a mere, 'blind' intuition into a cognised object. The imagination plays an essential role in cognition, albeit one that Kant never completely develops. In the *Critique of Pure Reason*, he refers to it as "the faculty for representing an object even without its presence in intuition" and explains that it constitutes a sort of bridge between concepts and objects by creating spatio-temporal interpretations of the categories (Kant, 1998 [1787], pp. B151, B176/A137–B187/A147). With this in mind, it seems promising to take the word 'imagination' very literally and to think of it as the image-producing faculty of the mind (Brook and Wuerth, 2004). This interpretation is supported by Kant expressing, at the very beginning of the *Analytic of the Beautiful*, that it is the imagination that relates representations to the subject, "perhaps combined with the understanding" (p. 5: 203).

Compared to their respective roles in cognition, the significance of the two faculties is inverted. Pure aesthetic judgements cannot be concerned with the cognised object as cognised, but they cannot be concerned with an immediate pleasure of the senses either, as this could result in no more than the gratification of a desire (p. 5: 206–5: 209). According to this essay's interpretation of Kant, pleasure that lets us properly judge something as beautiful can therefore only be experienced as the result of a process that allows for the exploration of the form of an object without the need to (accurately) cognise it. The understanding provides access to the form; the imagination provides the subject with the representation. The role of the understanding is not as extended here as in cognition since the application of concepts is not part of this interplay, and the role of the imagination is more pronounced than in cognition since the subject's reaction relies on the image produced in the absence of pre-existing categories.

But there is more to the relationship between these two faculties in the aesthetic experience of something as beautiful. In the absence of

predetermined concepts, they harmonise uniquely in what Kant calls 'free play' (p. 5: 217). It seems that the interest which is elicited by the contemplation of an object that we can properly judge as beautiful leads to a particular way of engaging with the object, in which the imagination and the understanding affect each other in a sort of chain reaction, animating each other reciprocally (p. 5: 287). For as long as the subject is disinterestedly interested in the representation as relayed to it by the imagination, the object's form will continuously be taken in by the understanding. Vice versa, as long as the understanding relates the object's form to the representation, a self-sustaining disinterested interest is produced. This results in an exploration of the properly beautiful object that goes on for as long as the object is attentively observed and is the ground on which disinterested pleasure emerges (p. 5: 222).

The relation of the imagination and the understanding in their 'free play' (insofar as it results in pleasure), and the way the faculties relate to one another as a condition of cognition in general, are both referred to, by Kant as well as commentators, as 'harmonious'. One might therefore object that the way in which the faculties interact in cognition and aesthetic contemplation must be similar. However, such a conclusion would result from a misunderstanding of how this harmony relates to the free play of the faculties (cf. Allison, 2001, p. 117). While is it a subjective condition of experience in general that the faculties act in accordance with one another, they do not interact in this back and forth 'play' in cognition. 'Free play' is unique to an aesthetic engagement with objects and cannot possibly emerge from cognition, as cognition is a mental act that produces transcendentally logical judgements (those that judge which categories to apply), and can therefore not cause an indefinite perpetuation of the state of judging in which the faculties remain engaged in the representation of the object.

Conclusion

After having shown how pure aesthetic judgements are related to nature in particular, without allowing for a strictly distinctive role of the imagination in the aesthetic appreciation of nature, I have argued that the role of the imagination in what Kant calls 'free play' is to contribute, as the primary factor, to the production of a pleasure that is qualitatively distinct from the pleasure that results from cognition. Given the interpretation of the imagination as an image-producing faculty, two aspects contributing to this qualitative difference were established. First, that it must play a much larger role in the aesthetic judgement of objects than in their cognition, and, second, that it both conditions and is conditioned by the understanding in the 'free play' that is

involved in pure aesthetic judgements.

NOTES

[1] This is not to say that it is possible to judge anything that is not cognised and hence not experienced. At any point in time at which we can make an aesthetic judgement, cognition has already occurred/is currently occurring. My point is that pure aesthetic judgements are unaffected by (and do not affect) the way in which categories are applied to the object in cognition. For instance, being mistaken about what an object is does not affect the pure aesthetic judgement of the same.

Bibliography

Allison, Henry E. (2001). *Kant's Theory of Taste: A Reading of the Critique of Aesthetic Judgement*. Cambridge University Press.

Brook, A. and Wuerth, J. (2004). 'Kant's View of the Mind and Consciousness of Self'. In: Zalta, E. N. and Nodelman, U. (eds) *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition). Metaphysics Research Lab, Stanford University.

Budd, M. (2005). 'Aesthetics of Nature', in Levinson, J. (ed.) *The Oxford Handbook of Aesthetics*. Oxford University Press, pp. 117–135.

Guyer, Paul (1997). *Values of Beauty: Historical Essays in Aesthetics*. Cambridge University Press.

Kant, I. (1956 [1788]). *Critique of Practical Reason*. Translated by Lewis White Black. Indianapolis and New York: Bobbs-Merrill Company.

Kant, I. (1998 [1787]). *Critique of Pure Reason*. Translated by P. Guyer and A. W. Wood. Cambridge University Press.

Kant, I. (2000 [1790]). *Critique of the Power of Judgement*. Translated by P. Guyer and E. Matthews. Cambridge University Press.