

Whence correctness?

Jaroslav Peregrin

Abstract. We know that lots of things are correct. (To help people who need it is correct. To move the bishop diagonally when playing chess is correct. To add 7 to 5 with the result 12 is correct.) But where does the correctness come from? I argue that correctness is best seen as something we humans created in the process of forming our societies. This, admittedly, is a speculation; but aside of this, there are facts that are more than speculations. In particular, I argue that our correctness is based on normative attitudes, but we let these often determine only the *criteria* of correctness, letting the criteria determine what is correct independently of our will. Thus it holds both that correctness is wholly our creation and it is independent of us in the sense that things may be correct even if we do not know that they are (or we think that they are not).

Keywords. Correctness; normativity; normative attitudes; criterial ascent; practices

Normative attitudes

We know that lots of things are correct (and lots are incorrect). To help people who need it is correct. To move the bishop diagonally when playing chess is correct. To add 7 to 5 with the result 12 is correct. (This is not to say that things neatly divide into correct and incorrect - there even seem to be things that are correct and also not correct. To hit somebody's nose in a boxing match may be correct according to the rules of boxing, and still incorrect if it causes the person an excessive suffering.) But how come? Where does the correctness come from? There are lots of answers philosophers have supplied us since the beginnings of their discipline (and recently there are also others given by scientists). Here I want to probe one of such answers: Correctness is something we humans created in the process of forming our societies. But to make this answer truly interesting, it is necessary to add a lot of details¹.

Simon Blackburn (1984) considered (and duly rejected) what he called the democratic harmony theory of correctness: to be correct, according to this theory, is to be "in tune" with

¹ It is clear that correctness is not quite homogenous concept: the correctness of helping people is of a kind different from that of chess moves. But they have something in common - I think that the fact that we use the term *correctness* in both (and many other) cases is not just a coincidence. Hence I think that despite the heterogeneity there is a substantial core that I want to address.

others, to do "the same" as others. While the sense of "the same" may be rather complicated here (think about an orchestra all the members of which play the same melody despite doing more or less different things), let us concentrate on straightforward cases. To reach the result 12 when adding 7 and 5 is correct because this is what (plus/minus) everybody does.

It is quite clear that this cannot be a viable theory of correctness, for we may often do correct things despite the fact nobody does them. Helping people in need seems to stay correct even if nobody does it. But imagine, for a moment, what it would bring us if the theory of democratic harmony *were* viable: it would, in effect, explain correctness in purely naturalistic terms, as it would *de facto* reduce correctness to regularity. It would smoothly embed correctness into the natural order - and this would be nice². Hence though the democratic harmony is a non-starter, we should consider whether we could perhaps exploit some similar ideas.

What, then, is correctness if not democratic harmony? Elsewhere (Peregrin, 2014; Peregrin, 2024) I argued that correctness is inseparably connected with our tendency to assume the so called *normative attitudes*. This concept was, to my knowledge, first discussed by Hart (1961) and then it became one of the cornerstones of the theory of normativity of Brandom (1994). This is my source of inspiration; though in the end I do not quite follow Brandom's lead.

So what is a normative attitude? It is, to begin with, a practical pro- or con- attitude to human behavior. We tend to encourage some behavior and we tend to discourage other. When, for example, somebody helps me, I approve of it, when she kicks me, I disapprove. But, mind me, these are *not* attitudes that would, in my eyes, aspire to be called normative, because they just promote my well-being. Attitudes that are justly called normative should be assumed from a more impartial viewpoint.

In particular, the attitudes should be sensitive primarily to the kind of behavior, not to who are its protagonists. Thus a normative attitude should oppose beating as such, not beating just *me*. Hence normative attitudes are what Tomasello (2014) calls „agent-neutral“ or what Giromini (2023) calls „projective“. But we should note that despite the name, normative attitudes do not, by themselves, yet introduce any genuine normativity³. For this, they are still too parochial and too insular. A rudimentary normativity comes into being only upon their resonance across a society. Or at least this is what I suggest.

² At least from the viewpoint of us, *naturalists*, who would appreciate having one unified picture of nature and our place within it.

³ From this viewpoint, it might be better to call these attitudes *protonormative*.

So in a sense we are back to democratic harmony, only now applied to normative attitudes. What I claim, then, is that something may become genuinely *correct* if it is targeted by positive normative attitudes resonating across the relevant society. I think this is the most primitive form of correctness from which, I am going to argue, other forms of correctness derive. (Also I am convinced that the existence of such a resonance can be seen as the existence of what can be called implicit rule - but I am not going to elaborate on this here⁴.) And there *must* be other forms of correctness; for there are obviously things that are correct despite being not generally approved of; and vice versa.

Practices

Hence though I claim that there is no correctness following the democratic harmony model (no correctness induced by mere regularity of behavior), there are correctnesses that are induced by the resonance of normative attitudes (i.e. induced by regularity of “metabehavior” - behavior targeting behavior). Hence any systematic interaction that may aspire to be a breeding ground of normativity must have the “metalevel” of evaluation of what is going on on the object level, in particular of normative attitudes. Elsewhere I proposed to call such interactions “practices” (Peregrin, forthcoming-a; Peregrin, 2024, Chapter 11).

What is the correctness that is generated merely by resonating normative attitudes? Consider, for example, meanings of linguistic expressions. Why is it the case that it is correct, in English, to infer *Fido is an animal* from *Fido is a dog*? It is hard to find any other justification than that this is what speakers of English have converged to taking for correct⁵. Could we take it for incorrect? Of course (only that then the words *dog* and/or *animal* would no longer mean what they mean now). There is nothing concerning the peculiar kinds of sounds constituting the English sentences that would sustain this correctness; the fact that these very types of sounds support this very inference can be hardly a matter of anything else than the attitudes of speakers⁶.

Or take greetings. Why it is correct to greet somebody at some place in some way (e.g. to say *Good morning* to a neighbor we meet leaving our house), while it is incorrect to greet somebody else at some other place in some other way (e.g. to say *Hey, dude!* to an old lady

⁴ See Peregrin (2014, sec. 4.1), Peregrin (2021).

⁵ Somebody may want to argue that the justification consists in the fact that the sentences mean what they do in English. I do not see the inference as separable from the meaning. Anyway, if somebody is willing to promote the separation, then the question will be that of the justification of the sentences having the very meaning they do.

⁶ Elsewhere (Peregrin, 2010; Peregrin, 2024) I called this kind of correctness *Protagorean*.

sitting on a park bench)? The answer, it seems, will, in the simplest case, be that this is what is felt like correct (and is buttressed by normative attitudes) in the relevant community.

Hence the idea is that resonance of normative attitudes generates the most rudimentary form of correctness. At the same time, however, this form of correctness cannot be – by far – the only one. Hence this basic form of correctness must have spawned some much more complex forms - perhaps all its other forms. How could that happen? My thesis, in a nutshell, is that this could happen by moving from correctness to its criteria. But before we elaborate on this, we must pause to consider whether it is reasonable to see the entire correctness as a human product.

"Unhuman" correctness?

Can correctness depend on us in this complete manner? It would seem that there are things that are correct even if nobody takes them for correct (think about complicated mathematical calculations), and that hence correctness, in this sense, is an *objective* matter. Hence is there not at least some correctness that is independent of us, that is just objectively here without our contribution? Let us consider some possible cases of such "unhuman" correctness.

When I ride a bike, I may do it correctly or incorrectly. Sometimes this correctness may have to do with our, human standards, but sometimes what we mean by "riding correctly" is just riding so that one does not fall down. And this does not have to do with our criteria or attitudes - nature itself decides who rides correctly in this sense and who does not. So is this not a case of "unhuman" correctness?

The problem, as I see it, is that this kind of *instrumental* correctness is only correctness in an attenuated sense of the word. For here "riding correctly" means "riding so as not to fall down" and this latter phrase does not seem to involve any normativity. Hence it seems that here "correctly" may be only a stylistic element that can be done away with. More generally, "You should (or it is correct to) do A to achieve B" is a stylistic variant of "If you do not do A, you will not achieve B".

Such instrumental correctness, it may seem, would be transformed into a correctness of a heavier caliber if there were goals that were in some sense absolute, that we could not fail having. Some authors think that this is the case of correctness instituted by evolution⁷. Here something is "correct" when it is in accordance with why it was selected for during evolution.

⁷ Most famously Millikan (1984; 2004).

The heart, to use the notorious example, works correctly if it pumps blood, for pumping blood was that for which it was selected.

Still, I hold this for an attenuated sense of "correct": it can be paraphrased away in terms of the talk about evolution. Here, something "is correct" is a shortcut for it "has been selected for by evolution" - viz. a purely descriptive statement. And I do not think that genuine "normatives", i.e. statements to the effect that something is correct or that it should be done, are (purely) descriptive. They are means of endorsement, of claiming allegiance to a norm (Peregrin, 2016; Peregrin, 2024, Chapter 6).

There is also a modification of this view, according to which some things are correct for us simply because we are *organism* and more things because we are the kind of organism we are. As Midgley (1981, p. 54) puts it: "You cannot have a plant or animal without certain quite definite things being good or bad for it." Hence something is good for us because it is a goal we cannot fail having - like "life itself"⁸. Sayer (2019, p. 262) writes: "Values and valuation in an intransitive sense are embedded in the way we are, in life itself, in the bodily strivings, appetites and aversions that keep us alive, some of which we interpret and evaluate transitively. There is an important sense in which life itself is normative, and what we subjectively regard as good or bad is partly an interpretation of that intransitive normativity."

I do not agree with this. It seems to me that we humans have reached the stage where we choose our values ourselves. True, some of them - like life - are accepted, as it were, by default; but even those may be rejected. (Is not Socrates valued despite - or even because - the fact that he drank the hemlock?) Therefore, I am convinced that correctness worth the name is our human product.

True, this may be seen as partly a terminological matter. Why not talk about the correct functioning of the heart? - This does not seem to be anything unusual. Well and good, but then we must explain whether this is an ambiguity of "correct", or whether there is a continuity between the different senses. And what I see is *discontinuity*: at some point of our evolution we became able to choose our values and thereby institute our correctness. True, we may expect that the values we choose normally chime with those established by evolution, or inherent to "life itself", but this is in no way a conceptual matter. I think that our correctness is *sui generis* and as such it should be distinguished from other phenomena in connection with which we may talk about correctness - and therefore I propose to use the term only for this our correctness.

⁸ This standpoint is often associate with neo-Aristotelianism (Rapp, 2020).

Criteria

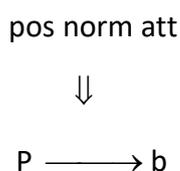
But if it is us who is the source of any correctness, how is it possible that correctness is often an objective matter, i.e. that we can be all mistaken w.r.t. what is correct? Here it is I think that we must enter the concept of *criterion*.

What, in general, is a criterion? The term is standardly explained as a characterizing mark or trait, usually some observable trait according to which we can say that the possessor of the trait is of some (not so easily observable) kind or has some (not easily observable) property. More generally, a criterion may be seen as a way of finding out that something has a property or is of a kind. Oversimplifying (very much) we can say that having gills is a criterion of being a fish, having the appropriate diploma is the criterion of being a medical doctor, and making a reliable scale show "5 kg" when positioned on it is the criterion of weighting 5 kg.

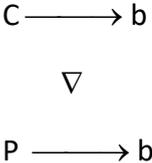
Now a criterion of a property or a kind is itself a property. *Having gills* is a property of carps, but not of whales. Moreover, a property may be linked to a property as being its criterion – a property C may be a criterion of another property P. Thus, *having gills* is a criterion of *being a fish*, while *having lungs* is not. And once it is settled what does have this property (i.e. the property of being the criterion of something), it is thereby settled what has the original property (the one of which it is the criterion). Once it is settled that *having gills* is a criterion of *being a fish*, it is settled that that this animal is a fish (as it has gills), whereas that is not (as it rather has lungs).

Note that the establishment of a criterion of having a property may bring about not only easier identification of the property, but also its precise delimitation. Once it is settled that the criterion of fishhood is having gills, there is no room for dispute whether whales are fish.

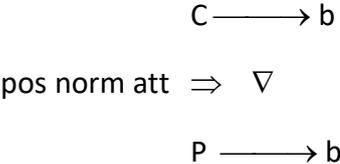
Now being correct is also a property (primarily of kinds of human behavior); and hence there may be its criterion. From what we said it might be concluded that in the simplest cases a kind of criterion may consist in the positive normative attitudes of a great majority of members of the relevant community. Diagrammatically, a kind of behavior, b, is correct (*viz.* has the property P of correctness) if the normative attitudes towards it are overall correct across the relevant society:



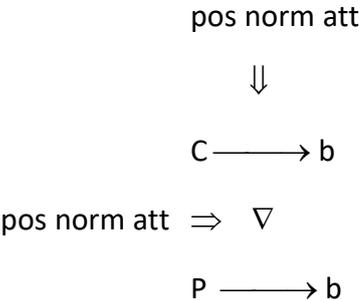
However, it is conceivable that there is a different criterion of something being correct. It is, for example, conceivable that something is a correct greeting iff it accords with an “unwritten tradition”, or that something is correctly called a *fish* iff it has gills. Diagrammatically, a kind of behavior is correct if fulfills a corresponding criterion:



How does this come about? Well, it must be clearly established by us, be it by a conscious and overt decision, or as a matter of a covert development. It is us who come to treat a property as a criterion of another property. And obviously it has happened that we have come to determine some correctnesses not directly by a resonance of normative attitudes, but by establishing its criterion. It is the resonance of our normative attitudes that determines that having gills is the criterion of fishhood; the resonance by itself no longer determining whether a particular animal is a fish. (We may all think that it is a fish despite the fact this is not true, that it does not have gills.) Diagrammatically, that C is a criterion of P (*viz.* that having C guarantees having P) is determined by the resonance of our normative attitudes:



Moreover, the point of the fact that P is grounded in C is in that C is in some sense more obvious than P, that that having C may be, in the simple cases, a matter of resonating positive normative attitudes.



Note also that the retreat from properties to their criteria is connected with an incipient form of what Brandom (1994) calls "the game of giving and asking for reasons". Once a concept is mediated by a criterion, the criterion can be used to produce a *reason* of the property. "Why is this fish-like animal not a fish?" "Well, it may look like fish, but it does not have gills!"

Criteria ascent

Imagine something is correct iff the bulk of the relevant community takes it for correct, if they assume positive normative attitudes towards it. Consider calling something *red*. It is red insofar as it is what the speakers of English call red⁹. But consider calling something a fish - we saw that here there may be an articulable reason, e.g. that the animal has gills. And here the agreement of a majority of speakers of English concerning an individual animal may not be a reliable lead - even a majority can happen be convinced that an animal, which *de facto* does not have gills, is really a fish. In such case, the natural outcome is that though the majority no longer decides whether a given animal is a fish, at least it decides what does it take to be a fish - what is the criterion of fishhood.

Imagine that an item has a property C just in case it is taken to have it - a great majority of people assumes positive normative attitudes to ascribing C to it. The attitudes thus decide whether the item has C or not. But imagine further that C is a criterion of P: C is a way of determining whether something has P. Then the normative attitudes not only determine what has C (in the direct way, i.e. by the majority vote), but also what has P (in an indirect way).

I think that moving from an agreement whether something has a property to the agreement on how to determine whether it has the property is a natural retreat strategy, analogous to the "semantic ascent" of Quine (1960, sec. 56). The ascent, according to Quine, is a move from "talk of objects to talk of words", it "carries the discussion into a domain where both parties are better agreed on the objects (viz., words) and on the main terms concerning them" (pp. 250-1). What we are talking about is close to a generalization of this move: again we move the discussion from disputing whether something has a property to disputing of what does it take to have the property. Let us call this *criteria ascent*.

The criteria ascent is a move from discussing whether something is P to discussing what does it take to be P. We discuss whether something is a fish or whether it is correct to greet somebody in some way, and move to discussing what does it take to be a fish (to have gills?) or what does it take to greet somebody correctly (to respect some "unwritten tradition" of the community in question?). We know that this may be useful within the kind of discussions we take part in now; but what I claim is that it might have been a move constitutive of this kind of discussions.

⁹ As Wittgenstein (1953, sec. 381) puts it: "How do I know that this colour is red?—It would be an answer to say: 'I have learnt English'."

What I claim is that taking something for correct if it is taken for correct by the bulk of the relevant community must soon turn out as inefficient: in many cases, no agreement is reached. The point is that we need to agree not only what words mean or how to greet whom, but rather also whether a specific (perhaps bizarre) animal is a fish or whether a specific (perhaps eccentric) display is a greeting. And here the criterial ascent is a natural way out.

It may be difficult to reach agreement about an item having a property, if the property is not something the item wears, as it were, on its sleeves. In some cases, it may be difficult to agree whether a given animal is a fish or whether a given display is a greeting. When this happens the criterial ascent is a plausible outcome. "If you say this is not a fish, then what does it take, for you, to be a fish?". And it may be satisfactory to reach an agreement on the criterion – to reconcile to the fact that we cannot always reach agreement w.r.t. individual items, but to know what we are talking about we must reach agreement w.r.t. the criterion.

Correctness inaugurated

In this way we can, I am convinced, reconcile the fact that correctness is wholly our creation with the fact that it is independent of us in the sense that things may be correct even if we do not know that they are (or we think that they are not). The reason is that we often set up correctness in terms of a criterion, and we are not always able to see whether the criterion applies.

It is tempting to use this to try to reconstruct the process in which correctness has found the way into our world and assumed the place it has in our developed societies. However, we must keep in mind that this cannot be more than a speculative tale (and it is not to be taken as a "just so story"). The beginning, according to the story told here, must have been the emergence of (rudimentary) normative attitudes – the emergence of the shared tendencies to prevent one's peers from doing certain things. (Maybe a "democratic harmony" played a role after all – maybe the things that were prevented were originally a matter of cutting down to size those who break ranks. But this would be a sheer speculation.) But the usefulness of correctness that is so directly linked to agreement is very limited – it could come to the whole fruition only after this link is loosened. And the productive way of loosening it is by way of the criterial ascent. Thus, according to this story, we have reached our current rich and multifaceted concept of correctness.

Aside of this, there are facts that are more than speculations. The fact is that the normativity that can be studied in human relationships is based on normative attitudes (Peregrin, forthcoming-b). The fact is also that we often determine the *criteria* of correctness, we let the criteria determine what is correct independently of our will. (This is most clearly seen in case

of language¹⁰.) And the fact also is that we humans have exacted from evolution the freedom to set our correctnesses more or less deliberately.

References

- Blackburn, S. (1984): *Spreading the Word*, Oxford: Clarendon Press.
- Brandom, R. (1994): *Making it explicit: Reasoning, representing, and discursive commitment*, Cambridge (Mass.): Harvard University Press.
- Frege, G. (1892): 'Über Sinn und Bedeutung'. *Zeitschrift Für Philosophie Und Philosophische Kritik* 100(1), 25–50.
- Giromini, J. (2023): 'The Logical Structure of Normative Attitudes'. *Philosophia* 51(3), 1271–1291.
- Hart, H. L. A. (1961): *The concept of law*, Oxford: Oxford University Press.
- Midgley, M. (1981): *Heart and Mind: The Varieties of Moral Experience (3rd ed.)*, London: Routledge.
- Millikan, R. G. (1984): *Language, Thought, and Other Biological Categories*, Cambridge (Mass.): MIT Press.
- Millikan, R. G. (2004): *Varieties of meaning*, Cambridge (Mass.): MIT press.
- Peregrin, J. (forthcoming-a): 'Inside Human Practices'. *Practices in Legal and Social Theories* (ed. W. Rzepiński et al.), London: Routledge.
- Peregrin, J. (forthcoming-b): 'Normativity between philosophy and science'. *Philosophical Psychology*.
- Peregrin, J. (2010): 'Language, the world and the nature of philosophy'. *The Philosophy of Richard Rorty* (ed. R. E. Auxier and L. E. Hahn), pp. 225–245, Open Court: La Salle (Ill.).
- Peregrin, J. (2014): *Inferentialism: why rules matter*, Basingstoke: Palgrave.
- Peregrin, J. (2016): 'Should one be a left or right Sellarsian? (And is there really such a choice?)'. *Metaphilosophy* 47(2), 251–263.
- Peregrin, J. (2021): 'Normative attitudes'. *The social institution of discursive norms* (ed. L. Townsend et al.), pp. 121–137, New York: Routledge.
- Peregrin, J. (2024): *Normative Species*, New York: Routledge.
- Quine, W. V. O. (1960): *Word and object*, Cambridge (Mass.): MIT Press.
- Rapp, C. (2020): 'Aristotle and Neo-Aristotelian Naturalism'. *Aristotelian Naturalism: A Research Companion* (ed. Hähnel, Martin), pp. 35–56, Cham: Springer International Publishing.
- Sayer, A. (2019): 'Normativity and naturalism as if nature mattered'. *Journal of Critical Realism* 18(3), 258–273.
- Tomasello, M. (2014): *A natural history of human thinking*, Cambridge (Mass.): Harvard University Press.

¹⁰ Frege (1892) famously urged the distinction between a sense [*Sinn*] and a meaning or a reference [*Bedeutung*]. His sense is close to meaning in the intuitive sense of the word; his meaning is what the word actually refers to. This two-layer conception of semantics has been integrated into most semantic theories since; and in many cases Fregean sense is just the criterion which makes it possible to determine the reference. The sense of the word *fish* does not involve individual animals, but gives you the criterion according to which you are able to tell whether a given animal is a fish. In this sense, whole language is a great sediment of criterial ascent.

Wittgenstein, L. (1953): *Philosophische Untersuchungen*, Oxford: Blackwell; English translation *Philosophical Investigations*, Oxford: Blackwell, 1953.