
3.8 Counterexamples

In everyday life, we often find ourselves asking whether we have done the right thing. Was it right to tell my mother that I never drink, or was it only a white lie? Was it right to have had all those drinks, or did I have such a good time that it does not matter if I woke up the neighbours? When doing philosophy, we are not concerned with only particular cases such as these. Our aim is also to discover more general truths, such as whether it can ever be right to tell a lie, or to find what it means for an act to be 'right' or 'wrong' at all.

This generality is what distinguishes philosophical questions from most ordinary questions. The answers philosophers put forward to their questions commonly involve generalizations and universals. They are statements that are supposed to apply to every relevantly similar instance of lying, not just the one in which you lied to your mother about your drinking. But it is because these answers are supposed to have universal or at least general application that individual cases become very important again, for an exceedingly powerful tool in philosophical thinking is the skill in deploying particular examples that undermine or at least qualify general claims. From a logical point of view universal claims (e.g. All X are Y) are extremely vulnerable to falsification because it only takes a single contrary instance to falsify them (Here's an X that is not Y). It is just this vulnerability that counterexamples exploit.

Good = pleasant example

For example, if we were to construct an argument to prove that 'good' acts are those that produce pleasure, we had better be sure that there are no instances in which an act could be deemed good even though it did not produce pleasure. If someone were to take us to task and produce such an instance, he or she would have cited what is called a 'counterexample'. The challenger might, for instance, suggest that giving money to charity is painful since it leaves you with less money for the finer things in life, yet few would suggest that donating a portion of your salary to the blind would not be a 'good' act. In this case, we will either have to renounce our hedonistic theory or else find a way for it to accommodate this counterexample.

We might, for example, reply that although you will experience pain as a result of your generosity, those who receive the donation will experience

pleasure. We will therein have made an important modification to our initial position (we might alternatively claim that it is a mere clarification): namely, that the pleasurable consequences that make an act 'good' do not necessarily have to be experienced by the act's agent.

In this way, counterexamples can perform the role of constructive criticism as well as being used to strike a theory dead. There was, of course, also nothing to stop us from *biting the bullet* and maintaining that giving money to charity is not a 'good' act at all. This may or may not get us very far. In the face of successive counterexamples and the theorist's responses to them, positions are honed until they are secure or else degraded until they are untenable.

Importance of the strange

It should be noted that counterexamples can involve some very strange hypothetical scenarios, but although such situations may be unlikely to occur in everyday life, this does not diminish their relevance in a philosophical argument. As a further counterexample to the hedonistic theory of goodness, it might be argued that there are individuals in the world – masochists – who achieve happiness by inflicting horrendous pain upon themselves. In their case, an act that resulted in their pleasure might not be regarded as good. Such individuals are rare, but if they do indeed achieve happiness through agonizing means, then they present just as pertinent a counterexample as the case of charitable donations. In short, a proposition or theory must be shown to survive even under outlandish conditions if it is to claim universal validity.

Limits of modification

So far so simple, but thinkers must also take care to preserve the essential nature of a position when subjecting it to trial by counterexample. Whether or not the essential nature of a position has been preserved when presented with a given modification or hypothetical scenario is often controversial. To take a famous example, the status of John Searle's so-called Chinese Room has been hotly debated. Supporters of 'strong artificial intelligence' maintain that a computer that passed the Turing Test (where computer responses could not be distinguished from those of a human, native-language user

in a blind test) would not merely be running a simulation of consciousness but would actually count as a full-blown mind possessed of cognitive states and the power of thought.

Against this argument, Searle constructed a counterexample. He imagined a room in which sat a person who understood not a single word of Chinese. Through a letter box the man receives questions written in Chinese characters and responds by looking them up in tables and passing back the symbols indicated by the table to be the appropriate answer. In essence, this is what a computer that apparently ‘understood’ Chinese would be doing, and, by that rationale, since the man in the room does not understand Chinese, neither would the computer. Both are functioning merely as mindless manipulators of symbols.

The ‘systems’ reply to the Chinese Room charges that Searle’s argument changes the nature of the putative possessor of any understanding. The man in the room may not understand Chinese, but the man *and* the tables within the room taken as a system do. It is the whole room that should be regarded as the language user if there is to be an accurate analogy for a symbol-processing computer. Just as we would not normally locate understanding in a special part of a Chinese speaker’s brain, neither should we expect understanding to reside in the computer’s CPU, for example. Though the whole, whether person or machine, may understand Chinese, any particular part of it might not. Since the strong artificial intelligence position is not committed to limiting the location of consciousness, it can be argued that Searle’s counterexample has altered the essence of the theory it was constructed to test. Defenders of Searle’s counterexample must show why this isn’t so.

SEE ALSO

- 1.8 Refutation
- 2.5 Anomalies and exceptions that prove the rule
- 2.9 Thought experiments

READING

- ★ John R. Searle, *Minds, Brains and Science* (1984)
- ★ Madsen Pirie, *How to Win Every Argument: The Use and Abuse of Logic* (2007)

3.9 Criteria

There's no great philosophical mystery about the meaning of criteria. A standard dictionary definition of a criterion is a 'standard by which something can be judged or decided'.

In this sense of the word, philosophy is full of criteria. Some are expressed in the form 'if and only if' (usually written 'iff') statements. So, if someone argues that a person has knowledge iff what she believes is justified and true, she is offering criteria for knowledge. In other words, something meets the standards of knowledge if it fulfils the conditions of being a justified, true belief.

In other contexts, the language of 'necessary and sufficient' conditions is used. In the above example, if the holding of a belief is justified and true, then all the conditions necessary and sufficient for knowledge are in place.

There is no good reason why, in standard English, either of the above should not be described as setting out the criteria for knowledge. But in philosophy, as in other disciplines, you should become sensitive to facts about usage. There are contexts where philosophers tend to talk about necessary and sufficient conditions rather than criteria and following them in doing this is advisable just because if everyone is using the same terms, everyone can feel more secure that he or she is actually talking about the same thing. Philosophers form a community of language users, and this community functions most smoothly if the same words are used in similar contexts.

There are dangers of ignoring this and seeing these conventions as little more than quaint pieces of academic etiquette. What you often find is that a perfectly normal word has become used in one corner of the discipline in a quite specific way. What then happens if you try to use it in another context is that confusion is created – are you using the word in its standard, English sense, or do you have the specialized usage in mind? Such is the case with 'criteria'. This word is now very much associated with the later work of Ludwig Wittgenstein (1889–1951).

Wittgenstein and criteria

Wittgenstein's work can be extremely gnomic, and sometimes it seems as though no two people agree on what it actually means. In broad terms, Wittgenstein made use of the idea of criteria for the meaning and use of words. For example, part of the criteria for the correct use of 'pain' is that

a person suffering pain behaves in a certain way: by showing distress, for example. The significance of using criteria here is that Wittgenstein is not saying that pain just *is* a certain form of behaviour, nor that such behaviour is a *sign* of pain, which is a private, subjective experience. The idea of criteria implies neither of those things – it merely specifies the standards for correctly using the word ‘pain’.

This, Wittgenstein believed, provided a way out of some old philosophical difficulties: How can we know that other people have minds? And how can I avoid solipsism – the idea that only I exist? These problems dissolve (rather than are solved) because the criteria for the correct use of words like ‘pain’ and ‘minds’ are behavioural and social – even though that does not mean that pain and minds *are* only behaviours. Hence the idea of criteria appears to be able to cope with the fact that the pains and pleasures of others are, in a sense, private, but that we have public rules for correctly using language about those aspects of our lives.

The state of Wittgenstein’s scholarship is such that none of the above should be treated as uncontroversial exegesis. Our key point is simply that the notion of criteria has both a special Wittgensteinian sense and an ordinary English sense. In the latter sense, ‘criteria’ seems to be a word that can be used across a wide range of philosophical discussions. But because of the former, it is wise in philosophical discussions to ration its usage, employing other words and phrases where they are available to avoid any confusion between the two. This is an important point, not just about criteria, but about the way in which apparently normal words get associated with particular philosophical positions. You need to be sensitive to this in order to express your arguments as clearly and unambiguously as possible.

SEE ALSO

- 3.2 Ambiguity
- 4.5 Conditional/biconditional
- 4.13 Necessary/sufficient

READING

Stanley Cavell, *The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy* (1979)