The Epistemology of Identity

Samuel Z. Elgin

It is my goal to lay the conceptual foundation for the epistemology of identity: to begin the investigation of how it is we come to know, justifiably infer, or reasonably believe that one thing is identical to another. My conclusion is a restricted form of skepticism; the only knowable identity claims are trivial. While I know that Hesperus is identical to Hesperus, that Aristotle is identical to Aristotle and that the property of being water is identical to the property of being water, I do not know that Hesperus is identical to Phosphorus, that Aristotle is identical to the pupil of Plato and the teacher of Alexander the Great, or that the property of being water is identical to the property of being the chemical compound $H_2O$. This paper is largely unconcerned with the metaphysics of identity. From the perspective of the world, as it were, substantive identity may well proliferate. But, if it does, it lies beyond our epistemic reach—a tantalizing relation that remains stubbornly and inexorably inaccessible.

I deny knowledge of substantive identity primarily because I do not see how it could arise. Initially plausible accounts face insurmountable obstacles. So, this paper is structured as a response to proposals. I discuss the possibility that our knowledge of identity is grounded in the identity of indiscernibles, spatiotemporal coincidence, the satisfaction of identity conditions, direct perception, and extant theories of evidence. These views fall into one of two camps: those that give rise to skepticism and those that are incorrect.

Yet hope remains. While the proposals I discuss to not yield knowledge of identity, they are inexhaustive. While I maintain that they are the initially plausible accounts, perhaps an initially implausible one will succeed while rivals have failed. And so, in some ways, this paper is as much a call to action as it is a philosophical argument. Although I see no resolution to the puzzle I raise, I deeply hope that one exists. Our epistemic lives would be bereft without knowledge of substantive identity.

Before considering candidate accounts, I offer a brief note on the distinction between trivial and substantive identity. Given that I allow for knowledge of trivial identity (but deny knowledge of substantive identity) this is an important distinction. It is tempting to suggest that a trivial identity claim is any that takes the form $a = a$—i.e., any in which

---

1Some terminological clarification: by an ‘identity claim’ I mean a sentence (or perhaps a proposition) of the form $a = b$ where both $a$ and $b$ are singular terms. The examples I use primarily involve cases when $a$ and $b$ are proper names, but I am also concerned with sentences of the form ‘$a = \text{The } F$’ and ‘The $F$ is the $G$,’ so definite descriptions figure centrally as well. And while I primarily discuss identity claims involving objects, I am also concerned with the identity of properties, propositions and relations.
the terms flanking the identity sign are themselves identical. However, for the purposes of this paper an epistemic, rather than a linguistic, characterization is preferable.

Identity is that relation everything stands in to itself and to nothing else. Knowing, as we do, the meaning of ‘identity,’ we conclude that Socrates is identical to Socrates, that the United States is identical to the United States, and that the property of being hydrogen is identical to the property of being hydrogen. That is to say, we sometimes obtain knowledge of identity by performing universal instantiation on the claim that everything is self-identical. A trivial identity claim is any that can be known via this method—simply through universal instantiation. A substantive identity claim, in contrast, is any that is not trivial. Hopefully, the presence of substantive identity claims is uncontroversial. However it is that chemists conclude that the property of being salt is identical to the property of being sodium chloride, it is not merely by appeal to universal instantiation. But if appealing to the meaning of ‘identity’ is not sufficient, what is? Under what conditions does someone know that one thing is identical to another?

The Identity of Indiscernibles and Inference from Identity Claims

An agent $S$ knows that $a = b$ just in case $S$ believes that $a = b$ on the basis of knowing that $a$ is indiscernible from $b$.

Perhaps the epistemology of identity can be accounted for by the identity of indiscernibles. Roughly, the thought is this: we epistemic agents regularly investigate the properties of things. We employ various empirical methods and inferential resources to uncover objects’ sizes, shapes, colors, masses, etc.. Sometimes, these methods reveal that objects have every property in common. An object $a$ bears property $F$ if and only if object $b$ bears property $F$. If, in this case, an agent were to conclude that object $a$ is identical to object $b$, she would do so justifiably (or, perhaps, the inference from ‘$a$ and $b$ share every property’ to ‘$a = b$’ preserves knowledge).

For example, a classicist might know through expert testimony that both Cicero and Tully were both famous Roman orators; she might know through original documents that both Cicero and Tully were born in the year 106 B.C.E.; she might know through history textbooks that Cicero and Tully were both sworn enemies of Mark Antony. Eventually, after a sustained investigation, she might conclude that Cicero and Tully have everything in common—that there is no property one has that the other lacks. On this basis, she might reasonably believe that Cicero is identical to Tully: the two men are one and the same.

This rough picture could be supplemented in various ways. Minimally, we need an account of how it is agents could know that objects bear all of the same properties. Humans have finite limitations, and have never determined all of the properties that something
bears. For example, our classicist friend is presumably unable to independently verify if both Cicero and Tully had an even number of hairs on their heads. Perhaps checking a wide, but inexhaustive, range of properties suffices, perhaps the best explanation of the fact that two objects share a great many properties is that they are identical, or perhaps some other story could be told. Further, some restriction must be placed on the kinds of properties that need be checked. If, in order to conclude that \( a = b \) one must first verify that both \( a \) and \( b \) bear the property being identical to \( a \), then in order to conclude that \( a \) and \( b \) are identical, one must first verify that they are identical. But however the account is to be augmented, the underlying thought remains the same: knowledge of identity arises from the identity of indiscernibles.

There is something appealing about this view. The identity of indiscernibles (and its converse) seem central to our conception of identity. It is no accident that identical objects bear all of the same properties. If someone were to conclude that objects \( a \) and \( b \) are distinct upon learning that their properties differ, they would do so justifiably. And so it is natural to suggest that if someone were to conclude that \( a \) and \( b \) aren’t distinct upon learning that their properties don’t differ, they would do so justifiably as well.

I suspect that astute philosophers already have objections in mind. In particular, this view faces a worry about Black (1952)-type worlds. Suppose there were a world consisting only of two perfectly homogeneous spheres. By stipulation, these spheres have every property in common—they are precisely the same size, shape, color, mass, etc. Often, this world is presented as a challenge to the identity of indiscernibles as a metaphysical view: i.e., as a counterexample to the claim that objects are identical just in case they bear all of the same properties. Surely, some argue, this situation is metaphysically possible. There is no conceptual challenge to imagining two qualitatively identical spheres. But if all qualitatively identical objects are numerically identical, then there would be one sphere, rather than two.\(^2\) This case problematizes not only the metaphysical version of the identity of indiscernibles, but the epistemic version as well.

Suppose someone were to observe such a world. Armed with acute attention and unparalleled perceptive powers, she recognized that the two spheres have every property in common. She realized that they are the same size, are equally massive, have the same material constitution, etc. And let us suppose that, after coming to know that the two spheres are perfect qualitative duplicates, she concluded that they are identical to one another. According to the epistemic view under consideration, she would thereby come to know that the spheres are identical. After all, she concluded that they are identical on the basis of knowing that they are indiscernible from one another. But this is false—the two spheres aren’t identical. There are two spheres—not one, and our imagined observer cannot know that distinct spheres are identical. This puzzle does not arise merely because knowledge is factive; her inference is presumably unjustified. She can plainly see two

\(^2\)For such discussions, see, e.g., Hacking (1975); Adams (1976); Hawley (2009). For defenses of the identity of indiscernibles, see e.g., Della Rocca (2005).
spheres before her very eyes, so how could her belief that they are one be justifiable?

This is a problem, but a potentially surmountable one. Some might maintain that a world consisting of qualitatively identical spheres is metaphysically impossible, and therefore irrelevant to the epistemology of identity. Even if it is metaphysically possible, further conditions could be added to correct for these kinds of cases. Perhaps we could add a ‘no-defeaters’ clause; if someone knows that $a$ and $b$ bear all of the same properties, the inference to the claim that they are identical generates knowledge as long as they lack epistemic defeaters. Because our imagined observer clearly sees two spheres, she has an epistemic defeater. Or, perhaps, we could introduce a condition involving spatiotemporal overlap. In order to justifiably conclude that $a = b$, one must not only discover that $a$ and $b$ share every property, but also that $a$ and $b$ perfectly overlap throughout space-time. Because our imagined observer knows perfectly well that the qualitatively identical spheres occupy different regions of space, her inference that they are identical does not generate knowledge.  

But although alterations may accommodate qualitatively identical situations, a deeper and more intractable problem lurks beneath the surface. Drawing out this problem is no simple matter, so I temporarily turn to an alternate (if rather silly) account that exhibits it more transparently.

Suppose that the transitivity of identity accounts for its epistemology: an agent $S$ knows that $a = b$ just in case she concludes that $a = b$ on the basis of knowing that $a = c$ and $b = c$. In order to know that Superman is identical to Clark Kent, Lois Lane must conclude that they are identical on the basis of knowing that there is some person $c$ such that Superman is identical to person $c$ and Clark Kent is identical to person $c$, and in order to know that Hesperus is identical to Phosphorus, an astronomer must conclude that they are identical on the basis of knowing that there is some planet $p$ such that Hesperus is identical to $p$ and Phosphorus is identical to $p$. This proposal isn’t simply the claim that agents can come to know identity claims by appealing to its transitivity; it is the further contention that the appeal to transitivity is exhaustive. It is the only way to come to know of substantive identity.

I doubt many readers find this account appealing, but the important point is this: if it is true, then the form of skepticism I advance obtains. This is because the transitive account gives rise to regress. In order to know that $a = b$ someone must know that there is a $c$ such that they know that $a = c$ and that $b = c$. But in order to know that $a = c$, someone must know that there is a $d$ such that they know that $a = d$ and $c = d$! And we’re off to the races—in order to know that one identity obtains someone must know infinitely many obtain. But because none of us do know infinitely many identity claims, no one knows any identity claims at all. Skepticism results.

Perhaps this argument preceded too quickly, so let us work through an example in more detail. Suppose Lois Lane, reflecting on numerous suspicions coincidences, concludes that

3I discuss criteria in terms of overlap in more depth in the following section.
Superman is identical to Clark Kent. Does she know that Superman is identical to Clark Kent? According to the transitive account, she knows that they are identical just in case she concludes that they are identical on the basis of knowing that there is someone that Superman is identical to and that Clark Kent is identical to. For example, if she knew that Superman is identical to the only alien on Earth and that Clark Kent is identical to the only alien on Earth, she could appeal to the transitivity of identity in order to know that Superman is identical to Clark Kent.

Does Lois Lane know that Superman is identical to the only alien on Earth? On the transitive account, she knows that they are identical just in case she concludes that they are identical on the basis of knowing that there is someone that Superman is identical to and that the only alien on Earth is identical to. This process continues without end; each step of the way, Lois is required to know yet another identity claim. And because this process continues without end (and because she does not know infinitely many identity claims) Lois does not know that Superman is identical to Clark Kent. This sort of regress affects knowledge of every substantive identity claim. And so, if the transitive account is correct, then no one knows substantive identity claims.

I suspect that many find the transitive account to be implausible—perhaps even ludicrous—partially for this reason. It engenders a regress that leads to skepticism about identity. It should be clear, however, that this regress could problematize other accounts. Any view on which knowledge of any one identity claim requires antecedent knowledge of another will lead to the very same regress. The only agents in a position to know that identity obtains would be those who know infinitely many identity claims. Because none of us do know infinitely many identity claims, these will all be views on which no one knows substantive identity claims.

I maintain that the identity of indiscernibles account face such a problem. Covertly—subtly—it builds in antecedent knowledge of identity. Knowledge of any one identity claim requires knowledge of another, and so it faces the same regress as the transitive account.

The starting point for this worry is Quine (1964)’s insight about quantification. He argued that sentences containing multiple occurrences of a variable within the scope of a quantifier presuppose the notion of identity. Consider the difference between ‘\(\exists xFx \land \exists xGx\)’ and ‘\(\exists x(Fx \land Gx)\)’. While the first sentence asserts that something is \(F\) and something (quite possibly something else) is \(G\), the second asserts that the very same thing is both \(F\) and \(G\).

---

4This argument could be framed equally well in terms of justified belief or possessing evidence, so while I focus on knowledge of identity here, the regress leads to a worry affecting the epistemology of identity more broadly.

5Others who provide similar arguments include Hawthorne (2003); McGinn (2000). For a recent critical discussion, see Burgess (2018). I will largely not engage with Burgess's argument, because his target differs from my own. He is primarily concerned with whether the concept of quantification involves the concept of identity (i.e., do invocations of the concept of quantifiers containing multiple occurrences of a variable invoke the concept of identity), while I am concerned with whether evidence of quantification requires evidence of identity.
and \( G \); that which is \( F \) is identical to that which is \( G \). Or consider the difference between \( \forall x Fx \leftrightarrow \forall x Gx \) and \( \forall x (Fx \leftrightarrow Gx) \). The first sentence claims that everything is \( F \) if and only if everything is \( G \); it is compatible with a situation in which an object is not \( F \) while some other object is not \( G \). In contrast, the second asserts that the very same objects that are \( F \) are \( G \); the objects which are \( F \) are identical to the objects which are \( G \). As Hawthorne says, “[w]ithout mastery of the concept of identity, it is not clear how we would understand the significance of the recurrence of a variable within the scope of a quantifier” (Hawthorne, 2003, pg. 100). I agree, and maintain that this has epistemic implications. In particular, someone who does not know that an object which is \( F \) is identical to an object which is \( G \) is not in an epistemic position to conclude that ‘\( D \, x \, p \, Fx \, ^D \, Gx \, q \)’. Similarly, someone who did not know that the objects which are \( F \) are identical to the objects which are \( G \) is not in an epistemic position to conclude \( \forall x (Fx \leftrightarrow Gx) \).

Although Quine and Hawthorne address first-order quantification, the point also applies to higher-order quantifiers. The sentence ‘\( \exists F(Fa \land Fb) \)’ differs from ‘\( \exists FFa \land \exists FFb \)’ in that the first presupposes that the very same property borne by both \( a \) and \( b \), while the second claims only that some property is borne by \( a \) while some property (quite possibly a different property) is borne by \( b \). The epistemic point remains as well; agents who do not know that the property borne by \( a \) is identical to the property borne by \( b \) are not in a position to know that \( \exists F(Fa \land Fb) \). That is to say, the only agents in a position to know that some property is borne both by \( a \) and \( b \) are those who know that the property borne by \( a \) is identical to the property borne by \( b \).

The identity of indiscernibles account states that an agent \( S \) knows that \( a = b \) just in case \( S \) believes that \( a = b \) on the basis of knowing that \( a \) is indiscernible from \( b \) — i.e., just in case \( S \) concludes that \( a = b \) on the basis of knowing \( \forall F(Fa \leftrightarrow Fb) \). Notably, this condition involves a variable that appears multiple times under the scope of a quantifier. In order to know that it obtains, agents must know that the properties borne by \( a \) are identical to the properties borne by \( b \). Those who do not know that the properties borne by \( a \) are identical to those borne by \( b \) are not in an epistemic position to know that \( \forall F(Fa \leftrightarrow Fb) \). Knowledge of this biconditional is itself necessary for knowing that \( a = b \) (at least on the present account), and so the only agents who know that \( a = b \) are those who know a further identity claim; they are those who know that the properties borne by \( a \) are identical to the properties borne by \( b \). And so, knowledge of one identity claim requires knowledge of others.

The same regress threatening the transitive account also affects the identity of indiscernibles. In order to know that \( a = b \), one must know that the properties borne by \( a \) are the same properties as those borne by \( b \), i.e., agents must know that the properties borne by these objects are themselves identical. In order to know that the properties borne by \( a \) are identical to the properties borne by \( b \), agents must know that their properties are identical. We are, once again, off to the races. And because the selection of \( a = b \) was arbitrary, knowledge of any one identity claim requires knowledge of others. Because no one knows infinitely many identity claims, no one knows any substantive identity claims.
at all.

I suspect that slight modifications to this view encounter the same obstacle. Even if an agent need not verify that \( a \) and \( b \) share every property in common (perhaps just checking a great many) she must still ensure that the properties she does check that are borne by \( a \) are identical to those properties borne by \( b \). The essential feature of the identity of indiscernibles is verifying that things bear the same properties; but in verifying that things bear the same properties, one must verify that the properties are themselves the same. I do not deny this account—perhaps it is the foundation of the epistemology of identity. But, if it is, then skepticism reigns.

**Spatiotemporal Overlap**

An agent \( S \) knows that \( a = b \) just in case \( S \) believes that \( a = b \) on the basis of knowing that \( a \) and \( b \) are spatiotemporally coincident.

How did the ancients discover that Hesperus is Phosphorus? The actual history is probably quite convoluted; scientists needed to develop an astronomical theory on which the same object could regularly appear in both the morning and evening sky, and some reason to predict that the object following Hesperus's trajectory would appear precisely where Phosphorus appears. I am no historian, and do not want to speculate about matters I know little about. Instead, I offer what is probably a fictionalized story. Astronomers, having developed the appropriate theory, recognized that Hesperus and Phosphorus occupied precisely the same region of space. Of course, they were presumably wrong about what this region of space was; the Babylonians who first realized that Hesperus is Phosphorus did not know that Venus orbits the sun. But although they were ignorance of what region of space Hesperus occupies, the knew it to be the same region of space Phosphorus occupies. Instead of believing that there are two planets that occupy exactly the same region of space, they concluded that the two planets are one and the same: Hesperus is identical to Phosphorus.

Perhaps the epistemology of identity can be understood along similar lines. After all, just as it is no accident that identical objects bear the same properties, so too it is no accident that identical objects occupy the very same region of space at the same time; Venus is not identical to an object located in a different place than Venus. Perhaps those who conclude that objects are identical on the basis of knowing that they are spatiotemporally coincident thereby come to know that those objects are identical. This view has attractive features. It easily accommodates the observations of Black-type worlds previously discussed. If someone were to observe two qualitatively identical spheres and conclude that they are identical, she would not count as knowing that they are identical. After all, she can plainly see that the spheres occupy different regions of space.

As with the identity of indiscernibles, refinements are needed. We humans have finite limitations and can neither check every region of space nor every region of time.
however the details are ultimately worked out, the underlying thought is that knowledge of identity arises from knowledge of spatiotemporal overlap.

Challenging cases quickly arise. Consider that classic example of the statue and the clay. Many hold that the two are perfectly coincident—the objects occupy the very same location in space. Nevertheless, we can be sure that the objects are distinct because they bear different properties. Minimally, their dispositional properties differ. A fragile statue is disposed to break when it is dropped, but a lump of clay would remain. This disposition might itself be understood in terms of essence. While the statue is essentially shaped thus-and-so, the clay is only accidentally shaped thus-and-so. So while an alteration of shape would destroy the statue, it would not destroy the clay out of which the statue is made.

Some discussions also allow for the statue and clay to have varied historically—the clay existed for some period of time before it was shaped into a statue. If so, the statue and the clay do not overlap throughout spacetime, they merely overlap through part of it. But we can consider versions of this case where both the clay and statue come into being and are destroyed simultaneously, and so vary in their dispositional properties but are always entirely coincident. Many philosophers thus contend that it is possible for two objects to occupy the same points throughout all of space-time. But why stop at two? How many objects are co-located with the statue?

The statue of David is currently located in Florence. It could be moved elsewhere—it might be displayed in an exhibition in Rome. Neither the statue nor the marble that compose it would be destroyed by this move; both only accidentally bear the property being located in Florence. But perhaps there is some other object which is destroyed. Perhaps something is coincident with David that is only accidentally composed of marble and only accidentally shaped thus-and-so, but is essentially located in Florence. This sort of object is destroyed when it is transported to another exhibition, but would not be destroyed if it were smashed. Of course, people rarely pay attention to this sort of object; it does not impinge much upon our lives. But some maintain that this reflects something about us, rather than the world. These philosophers (who are sometimes called ‘unlimited essentialists’) maintain that for anything that bears some collection of properties, there exists an object that bears each subset of those properties essentially and the rest accidentally. The primary motivation for unlimited essentialism is the difficulty in constructing a principled distinction between those properties an object bears essentially and the others accidentally. The what is it in virtue of that David essentially bears its shape but accidentally bears its location? Without a criterion in hand, it seems that what makes a property essential is an arbitrary feature of the world. Of course, we need not be unlimited essentialists in order

---

6 Those who defend objectual coincidence (in some form or other) include, but are not limited to Baker (1997, 2000); Fine (2003); Forbes (1987); Johnston (1992); Koslicki (2004); Kripke (1971); Lowe (1995); Oderberg (1996); Shoemaker (2003); Thomson (1983, 1998); Yablo (1987). For a defense of a conception of identity in terms of material constitution see e.g., Noonan (1993).

7 For defenses of unlimited essentialism, see Fine (1999); Johnston (2006); Koslicki (2008). For further discussion of unlimited essentialism, see Dasgupta (2018).
to maintain that the statue is distinct from the clay; perhaps there are two and only two objects at that location.

If there are two (or more) coincident objects, then challenges arise for accounts in terms of spatiotemporal overlap. Suppose someone came to know that the statue of David is co-located with a particular lump of marble, and concluded, on that basis, that the statue is identical to the lump. On the present account, such a person would thereby come to know that the statue is identical to the lump. But, surely, this is false; the statue isn’t identical to the lump, and so our reasoner cannot know that they are identical.

More mundane cases are also troubling. It is desirable for an epistemology of identity to account for the identity of both concrete and abstract objects: for it to provide conditions in which an agent knows that the property $F$ is identical to the property $G$ (for example). Accounts in terms of spatiotemporal overlap are ill-equipped for these kinds of cases, because abstract objects are not spatiotemporally located at all.

I suspect that modifications could account for these cases. Some philosophers might deny that the statue is distinct from the clay, and so contend that our imagined reasoner knows they are identical. Nominalists—who deny the existence of abstract objects—would presumably not object to accounts that do not provide conditions applicable to them.

I myself was briefly tempted by a hybrid view that incorporated elements from both the identity of indiscernibles and spatiotemporal overlap in order to accommodate these cases. I believed that an agent $S$ knows that $a = b$ just in case she concludes that they are identical on the basis of both knowing that $a$ and $b$ are indiscernible from one another and that they are spatiotemporally coincident. This, I thought, accommodated both the possibility of Black-type worlds and the distinction between the statue and the clay. Additionally, it might explain why it is our epistemic access between concrete an abstract objects differ. Abstract objects are not located in space or in time, so the condition of spatiotemporal coincidence is satisfied vacuously. In order to know that two abstract object objects are identical, one must merely verify that they bear the same properties. Concrete objects, in contrast, must also known to be located in the same region of spacetime.

I no longer believe this to be the case. The spatiotemporal coincidence account faces the very same regress affecting the identity of indiscernibles. In order to know that $a$ is identical to $b$, agents must antecedently know that the regions of spacetime they occupy are the same; i.e., that the regions of space-time they occupy are identical. Therefore, for an arbitrary identity claim $a = b$, knowledge that $a = b$ requires knowledge of some other identity claim. Because the selection of $a = b$ is arbitrary, knowledge of any identity claim requires knowledge of another. So the only agents in an epistemic position to know that an identity claim obtains are those who know that infinitely many identity claims. Because none of us know infinitely many identity claims, no one knows any identity claims at all.\(^8\) As with the identity of indiscernibles, I take no stand on whether accounts in

\(^8\)Note that this regress holds whether or not this account is supplemented by the identity of indiscernibles. In the unsupplemented form, the only identity claim that agents need antecedently know in order to know that $a = b$ is that the spacetime region occupied by $a$ is identical to the spacetime region occupied by $b$.  

9
terms of spatiotemporal overlap are correct; they may well furnish the foundation for the epistemology of identity. But, if they do, then skepticism reigns.

The Satisfaction of Identity Conditions

An agent S knows that \( a = b \) just in case S believes that \( a = b \) on the basis of knowing that the identity conditions for \( a \) and \( b \) are satisfied.

Other examples may serve as better guides. Quite plausibly, I know that \{2\} is identical to \{the successor of 1\}. My knowledge, it seems, arises from my inference from the claim that the number 2 is identical to the successor of the number 1 and from the identity conditions of sets; i.e., the fact that sets with identical members are themselves identical. That is to say, I conclude that the sets are identical on the basis of knowing what the identity conditions for sets are and knowing that these conditions are satisfied. Perhaps all knowledge of identity is similar. Perhaps knowledge of identity arises from the knowledge of identity conditions.

This suggestion engenders skepticism for a far more banal reason than regress; we typically do not know the identity conditions of things. Set theory is something of an exception; we know the identity conditions by fiat, simply because we stipulated those conditions when constructing the theory. Of course, knowledge of identity conditions alone does not itself furnish the resources for knowing that sets are identical. Knowledge that these conditions are satisfied is also required. In order to know that \{Hesperus\} is \{Phosphorus\}, for example, I must know that Hesperus is identical to Phosphorus. And so although I know what the identity conditions of sets are, I do not know that these conditions are satisfied unless I know further identity claims.

Other areas pose additional challenges. Take, for example, personal identity over time. Philosophers have plausibly devoted more attention to the identity conditions of persons than of anything else. There are a vast number of theories of what it takes for a person at one time to be identical to another.\(^9\) Perhaps an extant view is true; maybe Locke (1690) is correct that personal identity can be accounted for by memory. But even if Locke is correct, it is safe to assume that no one knows that Locke is correct. There is sufficient room for doubt about what personal identity consists of that no one has passed the threshold of knowledge. And if no one knows the identity conditions of persons, then no one can come to know that two persons are identical by knowing that the identity conditions are satisfied. After all, no one knows what these conditions are.\(^10\)

For the supplemented version, agents need also know that the properties borne by \( a \) are identical to the properties borne by \( b \).

\(^9\)This literature is so large, and I suspect that most readers are so familiar, that it is probably pointless to provide references. However, see, e.g., Locke (1690); Hume (1738); Reid (1785); Parfit (1971); Lewis (1976); Shoemaker (1984); Swinburne (1997).

\(^10\)Even if we were in a position to know what the identity conditions of persons are, other epistemic
In other areas, our ignorance is even more pervasive. We have devoted far less attention to (and so presumably do not know) the identity conditions for things like tables, planets or countries. On this account, in order to know that a cup of coffee is identical to the one I hold in my hand, I must know the identity conditions for a cup of coffee. But I do not know the identity conditions for a cup of coffee, so I do not know which cups of coffee are identical to my cup. Quite generally, we do not count as knowing what the identity conditions of things are. Because this account dictates that knowledge of identity requires knowledge of identity conditions, if it is true then no one knows substantive identity claims. And so, if knowledge of identity is grounded in knowledge of the satisfaction of identity conditions, skepticism reigns.

Knowledge Through Direct Perception

An agent S knows that \( a = b \) just in case S believes that \( a = b \) on the basis of directly perceiving that \( a = b \).

Perhaps we have unmediated access to identity.\(^{11}\) That is to say, perhaps we directly perceive a thing’s identity just as we directly perceive its color, size and shape. The epistemology of identity can be folded into the epistemology of perception; I know that a thing is identical to itself because I perceive that it is so.

The details of this suggestion depend upon the epistemology of perception. How do our perceptual experiences generate knowledge of the external world? Perhaps we draw conclusions from our sense-data.\(^{12}\) When we perceive a box is red, we possess a mental object (a sense datum) which is red, and conclude on that basis that the box is red. Or, perhaps, perceptions have finely-grained content (a perception \( p \) might have content \( c \) just in case those who perceive \( p \) non-inferentially conclude that \( c \)) and perceptions offer prima facie support for their content.\(^{13}\) And yet another route to the relevant type of skepticism would be to grant that we perceive identity, but to deny that perceptual experiences generate knowledge.\(^{14}\) But however the details are worked out, the underlying challenges would remain. If persons \( a \) and \( b \) are identical just in case either person \( a \) remembers acting as person \( b \) acted or person \( b \) remembers acting as person \( a \) acted, this account requires knowing what it is people remember. That is to say, if knowledge of identity arises from knowing that identity conditions are satisfied, and the identity conditions of persons are determined by their memories, then knowing that people are identical requires knowing what it is they remember. But, typically, we do not know what people remember. When I observe someone enter a room I have no idea what they remember and what they forget. And so even if we knew the identity conditions of persons (which we do not), we would often remain unable to know that people are identical.

\(^{11}\)My thanks to Ralph Wedgwood for this suggestion.
\(^{12}\)For defenses of this kind of view, see, e.g., Price (1932); Vogel (1990); Robinson (1994). For critiques, see, e.g., Berkeley (1710 (2008); Reid (1764 (1997).
\(^{13}\)For defenses of this kind of view, see, e.g., McDowell (1994); Brewer (1999). For critiques, see, e.g., Byrne (2005); Speaks (2005).
\(^{14}\)For discussions along these lines, see Davidson (1986), and, in places, Descartes (1673 (2017).
thought is that we directly perceive identity and so the epistemology of identity arises from
the epistemology of perception.

I maintain that this is false. To be clear, I grant that we could fold the epistemology
of identity into the epistemology of perception if we perceived identity. What I deny is
that we perceive identity. I am original doubtful because I know of know perceptual
mechanism that would allow us to perceive identity. Through sight, I perceive color;
through hearing, I perceive sound; through touch, I perceive shape. In contrast, I know of
no perceptual mechanism that allows agents to perceive identity. Rather, agents infer (often
unconsciously) that objects are identical on the basis of what they do perceive. However,
my initial doubts are no argument, and this proposal warrants a more substantive reply.

Suppose that I know two identical twins—Bert and Ernie—who are indistinguishable
to me; my perception of one of them is precisely the same as my perception of the other. Of
course, Bert and Ernie have no difficulty in distinguishing themselves ‘from the inside’—
Bert is not confused as to whether or not he is Ernie. But I lack this internal access, and
so I often am unable to tell them apart.

Suppose that I observe one of them enter the room. Because Bert and Ernie are
qualitative duplicates, my phenomenal experience would be the same whether or not the
person is either Bert or Ernie. For the sake of specificity, let us assume that the person
entering the room is Bert—not Ernie. The person who enters the room looks and sounds
precisely as Bert looks and sounds. Of course, he also looks and sounds exactly as Ernie
looks and sounds, for Bert and Ernie look and sound the same. What facts about identity
do I perceive?

There are four possibilities that could obtain:

1. I perceive that the person is identical to Bert and not identical to Ernie.
2. I perceive that the person is identical to Ernie and not identical to Bert.
3. I perceive that the person is both identical to Bert and identical to Ernie.
4. I neither perceive that the person is identical to Bert nor identical to Ernie.

Options 1 and 2 are nonstarters. Bert and Ernie are indistinguishable to me, so my
perception of one of them entering the room is the same as my perception of the other
entering. I lack a perceptual resource that allows me to determine that the person is
identical to Bert and not to Ernie, nor do I have a faulty mechanism which incorrectly
informs me that the person is identical to Ernie and not to Bert. There is nothing in my
perception that allows me to conclude that the person is identical to one rather than the
other.

Option 3 at least avoids the charge of arbitrary perceptions. If either person were to
enter the room, I would perceive that the person is identical both to Bert and identical
to Ernie. But this is where its advantages end. Option 3 entails that my perceptions are
incoherent with what I know to be true. If I perceive both that the person is identical to
Bert and identical to Ernie, then it follows from my perception entails that Bert is identical to Ernie. Because I know perfectly well that Bert is not identical to Ernie, my perceptual experience conflicts with what I know to be true.

But, surely, my perception isn’t inconsistent with what I know to be true. When Bert walks into the room I do not know whether the person is Bert or Ernie, but I have no doubt that the person is identical to one of them, rather than both. There is no epistemic need to re-weigh my evidence and determine whether Bert and Ernie after all. I don’t have evidence that they are identical. So I do not perceive that the person is both identical to Bert and identical to Ernie.

What remains is option 4. I neither perceive that the person is identical to Bert nor identical to Ernie. Because these are the only two people I could reasonably perceive this person to be identical to, I do not perceive identity. Of course, this case is somewhat idiosyncratic; not everyone has a phenomenally indistinguishable identical twin. However, my perceptual experience when Bert enters is presumably not dependent upon whether or not another phenomenally indistinguishable person exists. That is to say, it is not the case that whether or not I perceive Bert’s identity depends upon whether or not Ernie exists. And so I do not directly perceive facts about identity.

I thus deny that we directly perceive identity and, for this reason, also deny that the epistemology of identity is a species of the epistemology of perception.

**Extant Accounts of Knowledge and Evidence**

An agent S knows that \( a = b \) just in case S satisfies standard conditions for knowledge regarding the proposition that \( a = b \).

There is a suggestion that is boring, unoriginal, and initially plausible: the epistemology of identity can be straightforwardly folded into existing theories of knowledge, evidence and justification. There are several reasons to suspect that this is the case. First, it promises a unified theory of epistemology. We need not adopt a separate account for each type of proposition; rather a general account determines what agents know quite generally. Second, this explains why it is no one has bothered to discuss the epistemology of identity. There is no special problem for identity, the thought goes, we simply need to ‘plug in’ identity claims to existing theories.

A benefit of this suggestion is that it is easy to test. Accounts of knowledge, evidence and justification are readily available, so it is straightforward to select from the bevy of options. I do not canvass every epistemic theory here. This is partially due to restrictions on space, and partially because some accounts determine little (if anything) about what the epistemology of identity consists of. According to reliabilist theories of justification, for example, a belief is justified just in case the process by which it is formed is reliable—i.e., it leads to true beliefs sufficiently frequently.\(^{15}\) Absent an understanding of what it is that

---

\(^{15}\)For defenses of this sort of account of justification, see, e.g., Goldman (1979, 1986).
leads to the belief that one thing is identical to another, this says nothing in particular about the epistemology of identity. So it is difficult to determine what, precisely, justifies the inference to identity from this account alone.

**Epistemic Possibility**

Fortunately, other views are easier to evaluate. One such view relies upon the notion of *epistemic possibility*—a kind of modality relating to how the world might be for all that one knows. Epistemic modality is standardly contrasted with subjunctive modality—a kind not indexed to agents’ epistemic states. A prominent conception of epistemic modality concerns our ability to locate oneself within modal space.\(^{16}\) Roughly, the picture is this: there are a vast number of possible world. In some, Donald Trump is President, while in others Hillary Clinton is President; in some the United States never entered World War II, while in others it is currently engaged in World War III; in some intelligent life never evolved, while in others it is widespread throughout the universe. Presumably, no one has uniquely identified which possible world is actual, at least in a non-indexical way. While someone may truly say ‘This world is actual,’ she does not know whether or not she lives in a world with an even number of blades of grass. Nevertheless, some restrictions are at hand. I know that I am not in a world entirely devoid of intelligent life, for example, on the grounds that I exist.

Some maintain that evidence serves to eliminate candidate possible worlds. The more evidence I have, the narrower the scope of possible worlds that I might occupy. Thus, for example, Hawthorne (2004) states “It is possible that \(p\) for \(S\) at \(t\) (there is a chance that \(p\) for \(S\) at \(t\)) iff \(p\) is consistent with what \(S\) knows at \(t\)” (pg. 26. See also Hawthorne (2012)), and Stanley (2005) claims “It is possible \(A\) that \(p\) is true if what \(A\) knows does not, in a manner that is obvious to \(A\), entail not-\(p\)” (pg. 128).\(^{17}\)

Along these lines, we might say that \(e\) is *conclusive evidence* that \(p\) just in case \(p\) is epistemically necessary for all of those in possession of \(e\) (i.e., just in case for every agent in possession of \(e\), all epistemically possible worlds are worlds in which \(p\) is true). For example, my phenomenal experience is conclusive evidence for the existence of intelligent life, because there is no possible world in which someone possesses my phenomenal experience and intelligent life does not exist. Inconclusive evidence might be defined in various ways: we might say that \(e\) is inconclusive evidence that \(p\) just in case there are more possible worlds in which both \(e\) and \(p\) are true than worlds in which \(e\) is true and \(p\) is false, or perhaps that all (or most) of the relevantly close worlds in which \(e\) is true are also worlds in which \(p\) is also true, or that the counterfactuals ‘If \(e\) were true then \(p\) would be true’ and ‘If \(e\) were

\(^{16}\)There are other prominent accounts of epistemic modality. For example, (Hacking, 1967, pg. 153) defends the view that “it is possible that \(p\)” means that \(p\) is not easily known to be false, nor would practicable investigations establish that it is false.” For related views, see, e.g., Teller (1972); DeRose (1991); MacFarlane (2011).

\(^{17}\)For a more formal development of this sort of view, see, e.g., Yalcin (2007).
false then \( p \) would be false' hold. A newspaper report that Edmund Hillary climbed Mount Everest is inconclusive evidence that Hillary climbed Everest just in case the number of possible worlds in which the newspaper report exists and Hillary climbed Everest than possible worlds in which the report does not exist and Hillary did not climb Everest (or one of the other candidate theories of inconclusive evidence).

On this view, something constitutes evidence for an identity claim just in case it narrows which possible worlds we might occupy in the appropriate way. \( e \) is conclusive evidence that \( a = b \) just in case it epistemically necessary for all those in possession of \( e \) that \( a = b \)—i.e., all epistemically possible worlds are worlds in which \( a = b \) (and similarly so for inconclusive evidence). So, in which possible worlds does \( a = b \)?

Well, all of them or none of them, depending on whether ‘\( a = b \)’ is true. Since at least the 1970’s the received view has been that identity holds necessarily. If Hesperus is identical to Phosphorus in the actual world, then Hesperus is identical Phosphorus in every possible world; there is no possible way for them to be distinct. The necessity of identity receives support from multiple fronts; Marcus (1947) first offered a proof from modest modal assumptions, and Kripke (1980) provided additional linguistic support. If this is correct, then the set of worlds in which \( a = b \) is either the set of all possible worlds (if it is true) or else the null set (if it is false). On this view, everything constitutes evidence of (true) identity claims, as it is epistemically necessary for everyone that the identity holds, regardless of what evidence is at hand. This proposal thus fails the test of extensional adequacy, for it inaccurately entails that everything whatsoever is conclusive evidence of identity.

Perhaps some are tempted to appeal to metalinguistic ignorance in this kind of case.\(^{18}\) Although every possible world is one in which Hesperus is Phosphorus, not every world is one in which ‘Hesperus’ and ‘Phosphorus’ co-refer. Perhaps anything that narrows the possible worlds to ones in which ‘Hesperus’ and ‘Phosphorus’ denote one and the same object counts as evidence that Hesperus is Phosphorus.

This suggestion remains extensionally inadequate. Consider, for example, a person who has no knowledge of what either ‘Hesperus’ or ‘Phosphorus’ refer to—\( \text{they could denote planets, people, countries, etc.} \) Such a person might have it on good authority that ‘Hesperus’ and ‘Phosphorus’ co-refer—i.e., \( \text{they might hear expert testimony that the two names denote the same object, whatever that object is.} \) In this case, the person has evidence that ‘Hesperus’ refers to the same object as ‘Phosphorus,’ but she lacks evidence that Hesperus is Phosphorus.

So too, it seems, someone might have evidence that Hesperus is Phosphorus while lacking evidence that ‘Hesperus’ refers to the same object as ‘Phosphorus.’ Perhaps an astronomer, ignorant of the Ancient Greek names and out-of-touch with contemporary philosophy, has no idea whether or not ‘Hesperus’ and ‘Phosphorus’ co-denote. Such an astronomer could, with the appropriate astronomical evidence and theories, come to

---

\(^{18}\)See, e.g., Frege (1879); Stalnaker (1976) for this sort of appeal.
know that Hesperus is Phosphorus, without ever hearing either name. This astronomer
has evidence that Hesperus is Phosphorus, but lacks evidence that ‘Hesperus’ denotes the
same object as ‘Phosphorus.’ So it is possible to have evidence that ‘Hesperus’ co-refers
with ‘Phosphorus’ without having evidence that Hesperus is Phosphorus, and it is possible
to have evidence that Hesperus is Phosphorus without having evidence that ‘Hesperus’
co-refers with ‘Phosphorus.’

E=K

Perhaps this trouble is peculiar to this particular account of evidence. Perhaps other theo-
ries are better equipped to accommodate identity. And, perhaps, the ability to incorporate
the epistemology of identity is a mark in favor of alternate views.

To that end, I turn to Williamson (2000)’s recent proposal that evidence is knowledge.
He defends this view while outlining important theoretical roles knowledge can play even
if it is irreducible to anything more fundamental. In arguing that E=K, Williamson claims
that a person’s total evidence is identical to what they know. His argument is the following:

1. All evidence is propositional.
2. All propositional evidence is knowledge.
3. All knowledge is evidence.
4. Therefore, all and only evidence is knowledge.

This argument is incontrovertibly valid, but each of the premises is contentious. Many
prosecutors, for example, would enter a bloody knife as evidence in a murder trial, but a
bloody knife is not a proposition.

Williamson defends these premises in multiple ways. In support of premise 1, for
example, he notes that sentences of the form ‘A because B’ are ungrammatical if the terms
taking the place of A and B were not assertoric expressions; for example, the statement
‘Albania because ... ’ is ungrammatical. If evidence is to be understood as that which
underlies explanatory assertions, it may be that only truth-evaluable entities count as
evidence. In addition, particular evidence is sometimes said to be inconsistent with one
hypothesis or another, and it is unclear how something non-propositional is capable of
being either consistent or inconsistent with anything at all.19

The majority of Williamson’s discussion concerns agents’ total bodies of evidence. An
exhaustive theory of evidence surely requires more. In addition to accounting for the total
evidence someone has available, it ought, minimally, to determine when evidence lends

---

19The claim that E=K has come under attack on multiple fronts. For example, Brueckner (2005) argues
that Williamson cannot accommodate perceptual knowledge; I may know that a cup is red based on my
perception, but my perception is not a proposition, and Joyce (2004) argues that it cannot accommodate
the fact that the evidential status of many propositions is a matter of degree.
support to one claim rather than another. According to Williamson, my knowledge that
whales are mammals counts among my evidence, but which propositions does it support
in particular?

To this end, Williamson provides the following account:

\[ \text{EV: } e \text{ is evidence that } h \text{ for } S \text{ if and only if } S’s \text{ evidence includes } e \text{ and } P(h|e) > P(h). \]

Whether \( S’s \) evidence includes that \( e \) is determined by \( S’s \) knowledge, and whether or
not it is evidence for a particular proposition \( h \) is determined by whether the probability
of \( h \) given \( e \) is greater than the probability of \( h \) without \( e \).

This proposal faces the opposite problem plaguing epistemic possibility accounts. While
they entailed that everything is evidence of identity, Williamson’s proposal entails that
nothing is. On the standard view, identity claims are either necessarily true or necessarily
false. So the probability of any identity claim is either 1 or 0. According to standard
probability theory, it is impossible for any evidence to raise the probability of proposi-
tions whose initial probabilities are either. Nothing raises the probability of claims with a
probability of 1; that it the highest probability anything can possibly have. And nothing
impossible becomes more likely on the basis of new evidence; that which is impossible
remains impossible regardless of what evidence is obtained.

Evidence is defined in terms of knowledge which raises probability, and nothing raises
the probability of identity claims, so nothing is evidence of identity. This account thus fails
the test of extensional adequacy, for I assume that there is evidence of identity or other.
Accounts that entail that there is no evidence of identity (like the proposal that \( E=K \))
misdiagnose the relevant cases.

I have primarily addressed theories of evidence, but similar considerations apply to
modal accounts of knowledge. Nozick (1981), for example, defends an account in terms of
sensitivity; \( S \) knows that \( p \) just in case \( S’s \) belief that \( p \) is sensitive to the truth of \( p \)—i.e., if
and only if \( S \) would not believe that \( p \) if \( p \) were false.\(^{20}\) This too struggles to accommodate
necessary truths—it is difficult to interpret this condition when it is impossible for \( p \) to be
false. Similarly, Sosa (1999) defends an account of knowledge in terms of safety; \( S \) knows
that \( p \) just in case the \( S \) believes that \( p \) in the closest possible worlds in which \( p \) is true
and \( S \) does not believe that \( p \) in the closest possible worlds in which \( p \) is false.\(^{21}\) This
fairs slightly better than accounts in terms of sensitivity. Minimally, it requires that agents
believe identity claims in the closest possible worlds. Nevertheless, challenges remain.
Agents who stubbornly believe identity claims regardless of evidence count as knowing
identity on safety accounts. After all, if an identity claim holds at all then it holds in every
possible world. So the closest possible worlds in which an identity claim obtains are just
the closest possible worlds. So long as such agents hold fast to their belief in identity, they

\(^{20}\) For critiques of sensitivity, see Kripke (2011). For a recent defense, see Ichikawa (2011).

\(^{21}\) For critiques of safety see, e.g., Comestàna (2005).
will believe that it obtains in the closest possible worlds in which it obtains. And since there are no possible worlds in which the identity does not obtain, agents need not worry about disbelieving identities in the closest worlds in which the identity is false. But, it seems, steadfastly retaining belief in identity, regardless of the evidence, does not generate knowledge of identity.

The upshot is that accounts of epistemology which depend upon modal considerations are plagued by the necessity of identity. It is difficult for these accounts to accommodate propositions that are either necessarily true or necessarily false. Because identity holds necessarily (if it holds at all), it is difficult for these kinds of theories to account for the epistemology of identity. Of course, there are other theories without explicitly modal content. Whether any these view account for the epistemology of identity and can stave off skepticism is a subject for later discussion.

Conclusion

At the inception of the analytic tradition, there was a puzzle concerning identity. Philosophers came to realize that some identity claims are substantive while others are trivial. The claim that Hesperus is Phosphorus requires weighty empirical evidence, while the claim that Hesperus is Hesperus requires hardly any justification at all. What followed was a sustained discussion of the meanings of proper names, the development of a novel and meticulous philosophical methodology and the beginnings of the linguistic turn.

What did not follow was an investigation of the epistemology of identity. To the best of my knowledge, no papers were published on the conditions for an agent to know that one thing is identical to another; no conferences were held addressing what it is that justifies the inference to an identity claim; there were no books on what constitutes evidence of identity. Today, we remain nearly as ignorant of the epistemology of identity as we have ever been.

This paper has been an attempt to begin to rectify this historical oversight. I have canvassed the plausible accounts of the epistemology. Several (the identity of indiscernibles, spatiotemporal overlap and the satisfaction of identity conditions) give rise to skepticism; the only knowable identity claims are trivial. Another (direct perception) faces a challenge concerning observations of qualitative duplicates. And theories of knowledge and evidence that employ modal notions face challenges arising from the necessity of identity.

We are left in a bleak state. The most plausible candidate theories of the epistemology of identity generate a skeptical conclusion. If skepticism can be avoided, it is with an account I do not mention here. While this is the first systematic discussion of the epistemology of identity, I hope it will not be the last.
References


Descartes, René. 1673 (2017). “Meditations on First Philosophy.”


