

Relativistic Conceptions of Trustworthiness: Implications for the Trustworthy Status of National Identity Systems

Paul Smart, Wendy Hall, Michael Boniface

But he that is a friend to all men, is a friend to no man, and least of all to himself. For he must promise so much, that he cannot performe withall: and so breaking promise with some, he is trusted at length by none.

—Samuel Crook (2010/1658)

The term “Trustworthy National Identity System” suggests that trustworthiness is a particular feature of a national identity system (NIS), one that can be evaluated independently of the vagaries of time, place, circumstance, or the specific individuals (i.e., trustors) who might be inclined to place their trust in the NIS. This is what might be called an absolutist view of trustworthiness. The absolutist view allows for the idea that trustees (e.g., a specific NIS) can be trustworthy in a highly generic or simpliciter sense, which is to say that a trustee is equally trustworthy to all actual and potential trustors across a range of trust-related contexts.

Perhaps unsurprisingly, the absolutist view of trustworthiness has not gone unchallenged. For the most part, trust theorists have favoured a more relativistic conception of trustworthiness. According to this relativistic conception, trustworthiness is not invariant across trust-related contexts. Instead, trustworthiness is deemed to vary according to both the trustor and the matters with which the trustee is entrusted. This makes intuitive sense, since, in the case of human trustees, a single human individual may be trustworthy to some individuals but not others. Similarly, a single human individual may be trustworthy with regard to some matters (e.g., the provision of financial advice), but not other matters (e.g., looking after one’s children) (see Hardin, 2001).

These intuitions serve as the basis for a number of theoretical accounts of trustworthiness, all of which adopt a relativistic view of trustworthiness. These include so-called two-place accounts of trustworthiness, where trustworthiness is seen to be relative to particular trustors (e.g., Domenicucci & Holton, 2017), and three-place accounts of trustworthiness, where trustworthiness is seen to be relativized to particular trustors and particular spheres of activity (or domains of interaction) (e.g., Jones, 2012). Both two-place and three-place accounts of trustworthiness reject the idea that we can understand trustworthiness as being an invariant property of a specific individual (or, more generally, a trustee). This, it should be clear, has implications for the way we think about trustworthiness, as well as the way we develop and evaluate trustworthy systems. In particular, a relativistic conception suggests that it is a mistake to evaluate the trustworthiness of a trustee independently of the specific circumstances in which a trustworthy system will operate.

Relative to the contrast between relativistic and absolutist views, it is arguably the absolutist view that best serves the interests of those involved in the effort to develop trustworthy NIS. This is because the very notion of a trustworthy NIS implies that the relevant NIS has the capacity to be trustworthy in a highly generic sense, i.e., in a manner that ensures that it is *equally* trustworthy to all the citizens of a nation state. This idea is challenged by theoretical accounts that countenance a relativistic approach to trustworthiness. In particular, a relativistic view seems to question the basic possibility of trustworthy systems that are deployed at a national or supra-national scale. The reason

for thinking this is that it is hard to see how a single trustee (a NIS) could be equally trustworthy to trustors with competing or conflicting interests. This problem is arguably accentuated by the presence of multiple stakeholder groups (users, developers, politicians, commercial organizations, and other social actors) who may have different vested interests regarding the functional operation of a NIS.

In the present paper, we suggest that the absolutist view of trustworthiness is the appropriate view to adopt in respect of NIS. While this is counter to the bulk of the theoretical literature—much of which is wedded to a relativistic conception of trustworthiness—we suggest that it is possible for NISs to function in a manner that is consistent with the absolutist view of trustworthiness. Our approach is founded on two lines of argument: the first is to question the intuitive appeal of arguments that motivate relativistic conceptions of trustworthiness; the second is to highlight some of the practical steps that can be taken to ensure that NISs are trustworthy to multiple stakeholder groups.

In respect of the first point, we suggest that theoretical accounts of trustworthiness overlook an important feature of trust-related situations. This relates to the capacity of trustees to play an *active* role in ‘gating’ the placement of trust. If trustees take steps to prevent the placement of trust in situations where such trust is unwarranted, then a trustee cannot be accused of failing to fulfil trust (or betray the trustor) in situations where their trustworthiness might be called into question. Promissory obligations serve as a useful parallel here. One is a good promisor if one makes promises that one always keeps, but one is not a bad promisor for failing to keep promises that were never made. Being a good promisor and being trustworthy are, we suggest, somewhat similar in this respect.

Secondly, we highlight some of the practical steps that can be taken to ensure that NISs are uniformly (and equally) trustworthy to all those who rely on them. Here, we suggest that the functional remit of NISs ought to allow for the possibility of individualized forms of consent, control, and authorization, such that the functionality of a NIS is tailored to the interests, needs, and concerns of individual trustors. This calls for a degree of individual-level control regarding the operation of the system (e.g., identities only being disclosed to certain parties in certain situations), a degree of discretion regarding what the system will commit itself to doing in certain situations, and the implementation of safeguards that prevent the possibility of inadvertent harm to the trustor.

The upshot is a novel theoretical account that appeals to a variety of policy-related measures in ensuring a highly generic (absolutist) form of trustworthiness. Such an account, we suggest, moves us in the direction of a virtue-theoretic approach to technological trustworthiness (see Potter, 2002)—an approach that sees the pursuit of technological trustworthiness as tied to the development of technologies that surpass the limits of human virtue.

References

Crook, S. (2010/1658) *Ta Diapheronta, or, Divine Characters*. Kessinger Publishing, Whitefish, Montana, USA.

Domenicucci, J., & Holton, R. (2017) Trust as a Two-Place Relation. In P. Faulkner & T. Simpson (Eds.), *The Philosophy of Trust* (pp. 149–160). Oxford University Press, New York, New York, USA.

Hardin, R. (2001) Conceptions and Explanations of Trust. In K. S. Cook (Ed.), *Trust in Society* (pp. 3–39). Russell Sage Foundation, New York, New York, USA.

Jones, K. (2012) Trustworthiness. *Ethics*, 123(1), 61–85.

Potter, N. N. (2002) *How Can I be Trusted? A Virtue Theory of Trustworthiness*. Rowman & Littlefield, Lanham, Maryland, USA.