In the past four decades, transparency has gone from being primarily a clear plastic sheet used with overhead projectors to being primarily a quality of organisations. The pledge to provide (or, in corporate jargon, deliver) this quality has become an almost indispensable element of political speeches, earnings calls and product descriptions. The proliferation of digital technology and social networks has intensified this trend. This paper argues that our era’s fascination with transparency was prefigured by early modernity’s fascination with sincerity. A comparison of the two discourses shows that the functioning, purpose and benefits that 17th-century and 18th-century observers ascribed to sincerity are remarkably similar to those that 21st-century observers ascribe to transparency. This paper also interprets recent proposals for regulatory reform as digital-age refashionings of Immanuel Kant’s categorical imperative. Finally, it examines current speculations about the effect radical transparency is having on concepts of identity. Copyright © 2017 John Wiley & Sons, Ltd.
This paper argues that our fascination with transparency was prefigured by early modernity’s fascination with sincerity. A comparison of the two discourses shows that the functioning, purpose and benefits that 17th-century and 18th-century observers ascribed to sincerity are remarkably similar to those that 21st-century observers ascribe to transparency. Whether a particular person is sincere or a particular organisation is transparent is not at issue here. Instead, the binary distinctions sincerity/dissimulation and transparency/opacity are treated as schemas that early modern and present-day observers use to make sense of their world. The final section of the paper examines current speculations about transparency’s effect on concepts of identity.

FROM SOCIAL CONFORMITY TO SELF-CONFORMITY

The early modern discourse of sincerity was part of the wider discourse of upper-class interaction. Theorists believed that members of the upper class interacted with each other to create ‘agreeableness and diversion’ (Vaumorière, 1688, 7). For an activity dedicated to agreeableness and diversion, sincerity presents an irritant: ‘[I]t is to be considered how far sincerity should proceed; for I find nothing so ridiculous in Conversation, nor nothing more incommodious in the occasions of life, than for a man to speak whatever comes into his head’ (Vaumorière, 1688, 121). Experience shows that sincerity is not only incommodious; it often leads to insult and invective (Sacy, 1703, 8). Carried to an extreme, sincerity would render polite society impossible (Scudéry, 1683, 1: 169). Moreover, it is suspected that sincerity is rarely authentic and is in most cases merely a particularly subtle form of dissimulation (La Rochefoucauld, 1685, 131).

The advice offered by most 17th-century theorists is to refrain from saying what is on your mind and instead to conform with your interlocutors. Animals prefer to spend time with other animals of their own species, and it is a similar conformity that makes us enjoy each other’s company (Mére, 1677, 106). The comparison with the natural world is noteworthy, for in the 17th century the ‘only nature to be followed is that which will bring social acceptance’ (Krailsheimer, 1962, 83). In order to conform to a particular conversation or social situation, a member of the upper class must be able to judge ‘times and places and people’ (La Bruyère, 1687, 114; Vaumorière, 1688, 173; Barth, 1720, 3). Adapted from theories of rhetoric, this tripartite schema is a fundamental feature of the early modern literature on interacting in polite society. Because the circumstances of interactions vary, persons of quality are advised to develop a protean ability to dissemble. They should be learned with the learned, pious with the pious and sad with the sad (Barth, 1720, 112).

Not surprisingly, some members of the upper class find the norms of conformity and dissimulation stifling. They rebel by advocating authenticity and sincerity (Elias, 1994, 171). Their proclamations, like those of many rebels, have the air of a manifesto: ‘Sincerity must of necessity carry along with it all the Beauty of Truth, all the charms of Freedom, all the Sweetness of Confidence. It commonly produces a certain openness of heart, which appears in the Eyes, and renders the Physiognomy agreeable. Sincerity does not like Truth turn upon Words. All our Actions must also be sincere. It is ever an Enemy to Artifice and all Dissimulation; excessive Prudence is not in use with it. In a word, it is a beauty without paint, which fears not to be seen in the truest light, nor to be nearly observed’ (Scudéry, 1683, 1: 168–9). This plenitudinous definition, which is from a collection of model conversations, is met with several pages of objections from the interlocutors. But these objections, insists the sincerity advocate, eviscerate the concept, turning it into ‘a Quality that has nothing fix’d, which complys with Times, Occasions, and those to whom we speak’ (Scudéry, 1683, 1: 174). In short, attempts to circumscribe sincerity transform it into its opposite: social conformity with times (or more colloquially: moods), places and people.

In France, pro-sincerity sentiment is often a concomitant of social or political failure (Galle, 1985, 58–60). Disgrace and marginalization, it seems, afford greater freedom to experiment with
notions that contravene the norms of upper-class behaviour (Luhmann, 1980–95, 1: 121). In the early 18th century, a marginalized figure at the French court, the duchess of Orléans, repeatedly professes to be unable to conform and dissemble: ‘I don’t have wits enough to be able to conform to all the times. I’m sad or happy as the mood strikes me. The Lord made me too natural to be otherwise; ’I, who am so natural, believe that one should be what the time demands: happy when one feels good and sad when misfortune overwhelms one’ (Elisabeth Charlotte, 1867–81, 1: 28; 1908, 621; my translations). In seeking a discursive vehicle with which to explain her non-conformity, the duchess seems to adopt the paradigm of social conformity, its tautological structure and one of its three dimensions (time/mood). But she redefines this dimension as self-referential. The duchess is happy when her own mood is happy, not when the ambient mood is happy. She has replaced social conformity with self-conformity. And she has redefined ‘natural’ from that which brings social acceptance to that which prevents her from conforming with society and being accepted.

The duchess’s notion of involuntary non-conformity is analogous to other theoretical constructs of this era. For example, the early modern discourse of passionate love has been called an I-can’t-help-it ideology which is designed to enable lovers to circumvent the constraints placed on them by their matchmaking parents: ‘I can’t help it, father, I love her’ (Holmes, 1987, 41; Luhmann, 1982, 73). The duchess’s I-can’t-help-it ideology of non-conformity—she can’t help it, the Lord made her too natural to be otherwise—has a similar function: to enable her to circumvent the constraints placed on her by the norms of upper-class interaction.

FROM SINCERITY TO TRANSPARENCY, FROM FRIEND TO STAKEHOLDER

Sincerity, in addition to being a protest against the norms of early modern polite society, is increasingly seen as the solution to a problem. The problem, as described in 1717 by the French jurist and theorist Montesquieu in his essay In praise of sincerity (Éloge de la sincérité), is that people can only perceive their virtues and vices through the distorting veil of their amour propre. In all matters pertaining to themselves, people are unreliable witnesses and corrupted judges. This is why we have friends: to serve as each other’s guides and to tell each other what amour propre hides from us. In short, to tell each other the truth. Once our errors are pointed out to us, Montesquieu believes, we will endeavour to correct them because we cannot abide an obvious contradiction. The function of sincerity is therefore to make us better people. Insincere people who withhold the truth deprive us of a God-given right (Montesquieu, 1995, 7–10). Similarly, the duchess of Orléans states that the only service she can render to her friends is to encourage them to engage in self-examination and to help them rouse the God-given reason slumbering within them (Elisabeth Charlotte, 1867–81, 3: 145).

By the mid-17th century, the notion that we can either love God (which is good) or love ourselves (which is bad) has largely been replaced by the notion that self-love is at the heart of all human behaviour, including our love for God. Luhmann describes this as part of a broader shift from dualistic paradigms to self-referential paradigms (Luhmann, 1980–95, 1: 301–15). A famous example is René Descartes’s dictum, ‘I think therefore I am.’ For Descartes, thought confirms itself by means of its own operations, regardless of whether the contents of these operations are true or false, morally desirable or morally undesirable. Thought, in other words, confirms itself through self-reference. This, however, creates a potential problem of circularity. What is to prevent a self-referential subject from automatically referring everything back to itself, from automatically translating positive feedback into a positive self-image? The problem of circularity is exacerbated by the fact that upper-class interaction consists predominantly of expressions of mutual esteem and affection, which tends to make people see themselves as estimable and likeable (Nicole, 1671, 15).

This is where friends come in. Friends are sincere with each other. Indeed, sincerity is believed
to be the most essential element of a true friendship (Moser, 1763, 27). Our friends say things we do not want to hear, causing beneficial interruptions of our amour propre (Sacy, 1703, 84). In the language of systems theory, the function of friends in early modernity is to disrupt self-referential circularity (Luhmann, 1980–95, 2: 32).

Today, many people still consider this to be one of the functions of friends. Hollywood, for example, stages this function for dramatic and, often, comedic effect. The Sundance Kid tells Butch Cassidy, ‘It’s your “great” ideas that got us into this mess.’ In The Big Lebowski the Dude tells his bowling buddy, ‘Walter, I love you, but sooner or later you’re going to have to face the fact you’re a goddamn moron.’ Interestingly, Facebook friends rarely perform this function and instead generally engage in automatic affirmation and reciprocal compliments (‘U look gorgeous.’ ‘Thanks, hottie!’). Far from disrupting self-referential circularity, Facebook friends tend to encourage it (#mutualadmirationsociety).

The discourse of upper-class interaction stressed conformity and reciprocity as the most likely guarantors of pleasure and diversion. It attempted to create ‘a space where the Hobbesian war between different individuals ... would be suspended, at least for a time’ (Russo, 1997, 390). In this sense, Facebook may be the modern successor to early modern polite society: a space where the war between individuals—at home, in the workplace, on the road, in the grocery store—is mercifully suspended. Facebook is like a vast virtual Fontainebleau palace, where each day 1.7 billion courtiers post Facetuned selfies and greet their friends’ posts with an encouraging effusion of emoticons and exclamation points. Even when Facebook users post potentially controversial political news stories, which became common during and after the 2016 US presidential election, this too generally elicits automatic affirmation: ‘Yeah, Trump [Hillary] repulses me too.’ It takes a certain amount of civil courage to disrupt a Facebook friend’s political self-reference with a dissenting comment (although it does happen). For those who wish to engage in commentary beyond polite dissent (namely: snarkiness, public shaming and cyberbullying), other virtual forums exist: Twitter, cnn.com, Yik Yak, 4Chan and so forth.

In contemporary organisational theory, the functional equivalent of the friend is the stakeholder. Stakeholders disrupt an organisation’s self-referential circularity. An organisation’s most important disrupters, and in this sense its best friends, are its employees. They can enable the organisation, according to a theory developed by Karl E. Weick and Kathleen Sutcliffe, to be ‘mindful.’ A mindful organisation stops concentrating on ‘things that confirm your hunches, are pleasant, feel certain, seem factual, and explicit, and that others agree on.’ Instead, it concentrates on ‘things that disconfirm, are unpleasant, feel uncertain, seem possible, are implicit, and are contested’ (Weick et al., 2004, 17).

In Hollywood’s terms, a mindful organisation continually reminds itself that it may be a goddamn moron and continually reexamines its great ideas before they get it into a mess.

Regardless of the degree of mindfulness an organisation has achieved, it continually has its self-reference disrupted by outside stakeholders (customers, suppliers, investors, regulators, NGOs, concerned citizens, protesters). The number of stakeholders, their ability to observe and the size of the potential audience for their observations have increased dramatically over the past two decades. This new, radical transparency is described with metaphors drawn from the imaging and optical technology of the natural sciences: ‘Powerful institutional investors who now own or manage most wealth are developing x-ray vision. And in a world of instant communications, whistle-blowers, inquisitive media, and Googling, citizens and communities can easily put firms under the microscope’ (Tapscott and Williams, 2006, 22; my emphasis). The advent of X-ray machines and microscopes enabled doctors and scientists to see things previously invisible, from skeletal structure to cell structure. Similarly, the proliferation of information technology has enabled stakeholders to peer inside organisations and instantly communicate what they discover. This, observers believe, has consequences for organisations’ behaviour, consequences that echo those identified 350 years in the courtly advice literature.
Some of the early modern proponents of authenticity and sincerity advocated authenticity because of the difficulty of long maintaining an inauthentic performance. Sir Thomas Browne writes in the 1670s that ‘[m]en do not easily continue a counterfeiting Life, or dissemble unto Death. He who counterfeith, acts a part, and is as it were out of himself: which, if long, proves ... irksome’ (Browne, 1952, 279). A French advice book published in the next decade warns readers that esteem should be based on real virtue and that a reputation earned through dissembling will not last long: ‘Oh, how difficult is it to dissemble all the time!’ (Aubignac, 1687, 10). The discursive innovation in this advice is that it does not advocate the desired behaviour by invoking morality. Instead, it starts with a known characteristic of interaction—namely, that continuous, error-free dissembling is almost impossible—and from there argues for the advisability of the desired behaviour. In effect, it presents the desired behaviour as being forced on people by the exigencies of interaction.

Today, similar advice is being given to organisations. Executives are warned that ‘a lie won’t stay a secret for long in the consumer-powered Web’ and that, increasingly, no corporate misdeed goes unnoticed or unpublicised (Shiffman, 2008, 14; Clifford, 2002). Like the early modern advice, the advice for today’s executives leaves aside morality or at least concedes that morality alone is sometimes insufficient motivation. Instead, it argues that information technology, the internet and social networks make certain behaviours highly advisable: ‘The primary reason for not releasing a dangerous product ... should, of course, be a moral one. But every leader needs to keep in mind that the blogosphere is always there, waiting, watching, opining, and persuading’ (Bennis et al., 2008, 15).

DOES BEING IN THE BUFF ENCOURAGE ORGANISATIONS TO GET BUFF?

Confronted with the faults pointed out by our friends, Montesquieu optimistically believed, we will be unable to bear them and will seek to correct them. The duchess of Orléans exhorted her friends to self-examination in an effort to awaken the good sense slumbering within them. This, according to today’s observers, is also how transparency works: ‘Transparency networks encourage self-reflection as the firms caught up in them are continuously examined. Over time, firms develop greater awareness of their footprint on the world. This awareness, coupled with stakeholder scrutiny, drives firms to be more responsive to the norms and values acceptable to stakeholders’ (Tapscott and Williams, 2003, 25). Organizations are now ‘naked’, and, it is thought, the revelation of their imperfect bodies impels them to ‘get buff’ (Tapscott, 2012a).

Maybe it does. In 1989, the State of New York began requiring hospitals to disclose the death rates from coronary artery bypass graft procedure, a standardised and commonly performed operation. That year, hospitals’ death rates from this operation varied between 1% and 18%, indicating that a patient’s chances of survival varied significantly by hospital. Mandatory disclosure made this information available to patients, who, presumably, would avoid hospitals with high death rates. Hospitals responded to their sudden nakedness by reviewing their surgeons, clinical teams and recovery procedures in an effort to improve the quality of care. In the first 4 years of the programme, death rates fell by 41% statewide and have continued to fall since then (Makary, 2012, 34–47). Similarly, in the mid-1980s, the USA enacted a law that set no new emission thresholds but instead required companies to publish their emissions of toxic chemicals. Between 1988 and 1994, emissions of the chemicals covered by the law declined by 44% despite the fact that production increased by 18% (Florini, 1998, 58). Doubtless, there are other examples of the seemingly beneficial effect of transparency. Faith in the automaticity of this effect has led authors to declare in the subtitles of their books that transparency will ‘revolutionize health care’ and ‘revolutionize business’ (Makary, 2012; Tapscott and Ticoll, 2003). Maybe it will. For the purposes of this paper, however, it is enough to note that such assertions are increasingly common and increasingly plausible.

Copyright © 2017 John Wiley & Sons, Ltd.
The moral philosophy of Immanuel Kant exemplifies the shift towards generality. Published in 1785, Kant’s *Groundwork of the Metaphysics of Morals* (Grundlegung zur Metaphysik der Sitten) is the first moral philosophy to dispense with specific maxims. That is, moral judgments no longer have content, only form. ‘[T]here is an imperative, which ... is categorical. It doth not concern the matter of and what is to follow it, but the form and the principle, from which it issues itself, and the essential good of it consists in the mindedness [intention], let the consequence be what it will.’ (Kant, 1798–99–99, 1: 62; my emphasis). This categorical imperative is: ‘I ought never to act otherwise, than so, that I can be willing that my maxim shall become a universal law’ (Kant 1798–99–99, 1: 41). Kant does not tell you what you ought to do. Instead, he tells you how to determine whether what you intend to do is ethical. He does not give you a set of maxims but rather a maxim-checking app: could you wish that everyone would do what you are about to do? Take, for instance, a man who knows with certitude at the time he accepts a loan of money that he will never be able to repay it. Can he be willing for his maxim—‘When I am in want of money, I will borrow and promise to pay, though I know it will never be performed’—to become a universal law? The app says no: if everyone acted this way it would undermine the very notion of a promise and put an end to all lending (Kant 1798–99–99, 1: 71–72). According to Kant, the imperative to conduct such a universalisability check is the only imperative truly entitled to be called ‘moral.’

Today’s societies and markets are bewilderingly complex, as are the laws that regulate them. It is noteworthy that this has led some to propose a sort of categorical imperative for regulation: governments should stop focusing on content and instead insist on form. Instead of telling companies what to do, governments should confine themselves to requiring companies to be fully transparent. It is called ‘regulation by revelation’ (Florini, 1998, 53). The two laws described earlier work this way. Instead of setting new standards for heart operations or chemical emissions (content), they require disclosure of death rates and emissions (form). And they rely on the nakedness-induces-buffness mechanism to take care of the rest.

**IS AUDIENCE SEGMENTATION MERELY ERODING OR ENDING?**

Each of us has a range of identities. Which one we display—the knowledgeable sports fan, the gossipy neighbour, the decisive executive, the irate driver—varies by interlocutor and setting. In the words of philosopher William James, ‘a man has as many social selves as there are individuals who ... carry an image of him in their mind... We do not show ourselves to our children as to our club companions, to our customers, as to the laborers we employ, to our own masters and employers and to our intimate friends’ (James, 1890, 1: 294). Erving Goffman, who likens every-day interactions to performances between actors playing parts on a stage, calls this phenomenon ‘audience segregation.’ By practising it, ‘the individual ensures that those before whom he plays one of his parts will not be the same individuals before whom he plays a different part in a different setting.’ Interestingly, Goffman believes that the audience has as much interest in maintaining segregation as the performer does. ‘The audience can see a great saving of time and emotional energy in the right to treat the performer at occupational face value. Urban life would become unbearably sticky for some if every contact between two...
individuals entailed a sharing of personal trials, worries, and secrets’ (Goffman, 1959, 49).

Many UberX drivers appear not to have read Goffman. Although Uber instructs them to ‘avoid inappropriate topics of conversation’ (Snelson et al., 2016, 14), they typically do not avoid conversation. Indeed, they frequently initiate it, sometimes sharing what Goffman calls ‘personal trials, worries, and secrets.’ In October 2016, Britain’s Employment Tribunal ruled that Uber drivers in London are workers employed by Uber, not independent contractors using the Uber app to operate their own individual businesses (Snelson et al., 2016, 1). From a sociological perspective, however, one could argue that in a not insignificant way UberX drivers do not act like the employees of a transportation company because they do not abide by audience segregation. Instead of adhering to their ‘occupational face value,’ they play much the same part before customers that they play before friends. They are, in Goffman’s terms, ‘disrupting’ the traditional definition of the situation and imposing, or seeking to impose, a new definition (Goffman, 1959, 1–16). In Uber world, the drivers seem to contend, there is no longer a Driver and a Customer, just Hipster A and Hipster B, one of whom happens to be guiding the car (a definition supported by the fact that no money changes hands). Some customers may find the redefined situation, in Goffman’s words, ‘unbearably sticky’; others may find it refreshingly congenial.

The redefinition of the situation by some UberX drivers represents an incremental erosion of audience segregation in the interaction between driver and customer. Similarly, the audience, despite the advantages it derives from segregation, may itself seek to redefine the situation in a way that erodes segregation. For example, many customers at a tourist hotel in Cabo, Mexico, refused to treat Carlos, our beachfront waiter, at occupational face value. Instead, they insisted on treating him like a new-found friend. There was no longer a Waiter and a Customer, just Beachgoer A and Beachgoer B, one of whom happened to be handing out piña coladas (and because it was an all-inclusive hotel, no money changed hands here either). Many other examples of such incremental erosion could be adduced.

Some observers, however, think that audience segregation is being eroded across the board in far-reaching ways. Indeed, they believe that the instant availability of abundant information about nearly everyone is creating an unprecedented transparency that will soon render audience segregation impossible. One of these observers is Mark Zuckerberg, the CEO of Facebook: ‘You have one identity. The days of you having a different image for your work friends or coworkers and for the other people you know are probably coming to an end pretty quickly... [T]he level of transparency the world has now won’t support having two identities’ (Kirkpatrick, 2010, 199).

William James said that we have as many identities (‘social selves’) as we have groups of people who have a different image of us. Mark Zuckerberg says: one identity, one image. Similarly, the CEO of a company that manufactures all-natural snack bars segues from the authenticity afforded by his products’ transparent wrappers to the authenticity afforded by a single identity: ‘If you are authentic, you learn to move past the urge to be different things to different people’ (Lubetzky, 2015, 173). The schema single/multiple was part of the early modern discourse of authenticity as well: ‘Swim smoothly in the stream of thy Nature, and live but one Man. To single Hearts doubling is discruciating’ (Browne, 1952, 279). Inauthenticity is impracticable—Oh, how difficult is it to dissemble all the time!—so choose authenticity instead.

What Zuckerberg and others are diagnosing now is something different, however. According to them, authenticity is being forced on people not by the exigencies of interaction but by something outside interaction, namely: technology-driven radical transparency. The new situation calls for a new code of behavior, one dictated by the reality that we can never assume we are alone or unwatched’ (Bennis et al., 2008, 17–18). This too constitutes a sort of categorical imperative for the digital age: ‘I ought never to act otherwise, than so, that I can be willing that my action shall be captured on video and go viral.’

Someone who faced this choice nearly 40 years ago was Roddy, a character played by Harvey Keitel in Death Watch, a science fiction film by Bertrand Tavernier released in 1980. Roddy, who
works for a reality television programme, agrees to have miniature video cameras and transmitters surgically implanted behind his eyes (effectively making him the first owner of iGlasses). Everything he sees is simultaneously transmitted to a TV studio, where it appears on monitors in the production room and is recorded on tape. The cameras enable him to film a participant in the programme when she does not suspect it, thereby enhancing the programme’s reality. But before the programme enters production, Roddy visits his estranged wife. The two are about to have sex when Roddy realises that his view of their act will be transmitted to the studio for his boss and coworkers to see. Does Roddy apply the digital-age categorical imperative? Yes and no. Unwilling to share his own intimate moments with his coworkers, he leaves his wife’s house before anything happens. But this reticence in his own case does not prevent him from fulfilling his function as a surreptitious cameraman in the reality programme.

One of the consequences of functional differentiation is that each societal subsystem uses its own criteria to determine who may participate in it. Over time, participation tends towards full inclusion: universal suffrage, universal access to health care, universal access to education, and so forth. The subsystems have a high degree of indifference to each other. That is, our status in one subsystem generally does not affect our status in others. For example, having 12 children may make you strange in some people’s eyes, but it will not prevent you from being able to participate in the economy, receive healthcare, attend sporting events, vote or obtain legal counsel (Fuchs, 2000). In this way, functional differentiation itself fosters audience segregation.

At least it used to. The radical transparency diagnosed by Zuckerberg is already affecting hiring decisions. Depending on the source of the statistics, 35% or 70% of companies reject job applicants because of unflattering information they find in social networks (Kirkpatrick, 2010, 204; Tapscott, 2012b). Having risqué pictures on your Twitter or Facebook page may not prevent you from being able to vote or obtain legal counsel. But it may prevent you from getting the job you want. In this case, the functional subsystem is still using its own criteria (the binary distinction between suitable and unsuitable job candidates) to regulate participation. But when applying this distinction, it now has substantially more information to draw on. This new transparency lessens job applicants’ ability to ensure that the part they play during spring break in Florida or during vacation on Ibiza is segregated from the part they play during a job interview.

This is an example of how transparency has tangibly eroded audience segregation in one societal subsystem. Wikileaks may be another. During the 2016 US presidential election, the publication of hacked private emails made it harder for some Democratic party leaders to play a different role before voters than they played before each other (and behind each other’s back). There are likely other analogous examples.

Going forward, the broader empirical question is whether increasingly radical transparency will make the functional subsystems less indifferent to one another and whether it will reverse in any meaningful way the trend towards full inclusion. Meanwhile, we can observe how other observers deploy binary distinctions—transparent/opaque, authentic/inauthentic, single identity/multiple identities, real news/fake news—to try to make sense of it all.

REFERENCES

In this paper, the contemporaneity and chronology of source material are relevant. In seven cases where the year of the first edition deviates significantly from the year of the edition I consulted, I therefore used the year of the first edition in the page references in the body of the text. This enabled me generally to dispense with stage-setting (‘In a book published in 1660s,’ and so forth). The year of the edition I consulted is shown in parentheses at the end of the corresponding bibliographic entry.
