

UC Irvine

UC Irvine Previously Published Works

Title

Information ethics and the law of data representations

Permalink

<https://escholarship.org/uc/item/78g3x7mn>

Journal

Ethics and Information Technology, 10(2-3)

ISSN

1388-1957

Author

Burk, Dan L

Publication Date

2008-09-01

DOI

10.1007/s10676-008-9161-y

Peer reviewed

Information ethics and the law of data representations

Dan L. Burk

Law School, University of Minnesota, 229, 19th Avenue South, Minneapolis, MN 55455, USA
E-mail: burkx006@umn.edu

Abstract. The theories of information ethics articulated by Luciano Floridi and his collaborators have clear implications for law. Information law, including the law of privacy and of intellectual property, is especially likely to benefit from a coherent and comprehensive theory of information ethics. This article illustrates how information ethics might apply to legal doctrine, by examining legal questions related to the ownership and control of the personal data representations, including photographs, game avatars, and consumer profiles, that have become ubiquitous with the proliferation of information and communication technologies. Recent controversy over the control of player performance statistics in “fantasy” sports leagues provides a limiting case for the analysis. Such data representations will in many instances constitute the kind of personal data that information ethics asserts constitutes an information entity. Legal doctrine in some instances proves sympathetic to such an assertion, but remains largely inchoate as to which data might constitute a given information entity in a given instance. Neither is information ethics, in its current state of development, entirely helpful in answering this critical question. While information ethics holds some promise to bring coherence to this area of the law, further work articulating a richer theory of information ethics will be necessary before it can do so.

Key words: avatar, copyright, data representation, Floridi, information entity, information ethics, information law, intellectual property, law, privacy, right of publicity

Introduction

Over the past decade, Luciano Floridi and his collaborators have articulated an increasingly elaborate vision of information ethics, an ethical system based upon the primacy and moral worth of ordered informational structures (Floridi 1999, 2003; Floridi and Sanders 2001). Building from an established base in computer ethics, they have developed insights from information theory into a generalized framework for ethical analysis. Central to this system is the concept of the information entity, an autonomous information object inhabiting an infosphere comprising both tangible and intangible informational patterns. Since the Floridian infosphere encompasses the environment inhabited by information entities, it broadly includes informational interactions among material objects. Nonetheless, even though the world of atoms is awash in information, informational objects, and information entities, and information ethics has broad applicability to physical interactions, it is from the world of bits, from cyberspace if you will, that the theory arises, and from which many of its most compelling case studies arise.

Such a system of ethical principles has clear implications for legal doctrine. Law as a formalized and applied set of ethical practices is ultimately grounded in some framework of guiding principles; information ethics aspires to provide such principles. At its broadest expanse, information ethics potentially applies to essentially every area of the law, to all the rules concerned with arranging and ordering life in the world of atoms (Negroponte 1995, p. 237). But it is in the areas of information law, in the legal doctrines that are increasingly concerned with bits rather than with atoms, that information ethics may find its first and most straightforward application. Just as information ethics arises out of computer ethics, its application to the legal structures grows most naturally out of an application to information law.

Although he has said relatively little to date about the application of information ethics to information law, it seems clear that this connection has not escaped Floridi. His illustrations draw on the language of legal categories. He compares invasion of privacy to personal injury such as assault or kidnapping and rejects comparisons to

trespass (Floridi 2005, 2006). He has already begun some preliminary exploration of the relationship between information ethics and privacy, and by implication of the relationship between information ethics and the law of privacy (Floridi 2005). Such a relationship in turn implicates other areas of information law, most especially privacy's fraternal twin, the right of publicity, and an array of intellectual property regimes that might be applied to information objects.

Comprehensive mapping of information ethics to legal principles generally, or even to information law generally, will be an extensive undertaking. But we can begin such a project on a smaller scale, considering in a limited case the ways in which information ethics may be consonant with current information law, and the ways in which it might be dissonant. In this essay I trace the contours of a legal problem to which information ethics might be applied, cataloging the relevant legal doctrines, their ethical underpinnings, their compatibility with a theory of information ethics, and the prospects that information ethics might provide an ethical framework superior to that currently in place. My focus is on the kind of constitutive data that Floridi argues makes up an individual as an information entity, and most particularly on collections of personalized information that are intended to represent all or part of the individual. Such data representations might range from a photograph to a genomic database to a gaming avatar. I hope to show what account the law presently takes of such representations, and what account it might take of them with the application of information ethics.

My entry point into this investigation is the recent case considering the right of publicity claims for athlete performance statistics used in Internet "fantasy" sports competitions. This case allows me to introduce both the right of publicity and a limiting case for legal recognition of data representations. I sketch the relationships between rights of publicity, rights of privacy, and other bodies of intellectual property that might be applied to personal data representations. I then turn to the assignment of ownership implicit in various theories of information law, the ethical assumptions behind such assignments, and the structural gaps in those assumptions. Having identified certain deficiencies in the current legal framework, I ask whether information ethics might be able employed to better structure information law. I conclude that information ethics in its current form is not up to the task, but that a richer theory of information ethics holds promise for lending coherence to the law relevant to data representations.

The law of fantasy sports

Professional athletes and their affiliated teams have recently asserted proprietary control over the representation of the sports figures in on-line "fantasy" sports competitions. In such fantasy competitions, a competitor specifies on the Internet a virtual team of athletes, based upon statistical configurations from the real-world performance of physical sports players. Fantasy teams progress in virtual tournaments according to the corresponding real-world performance of the players represented by the statistical sets. In *C.B.C. Distribution and Marketing, Inc. v. Major League Baseball Mass Media* players and their teams lost a challenge to the use of these statistics, but the resulting opinion of the court provides an entry point into the law concerning data representations.

Although the case encompassed a variety of claims and counter claims, the court's holding was centered upon the plaintiff's right of publicity claims. In the United States, the right of publicity evolved, primarily at common law, with some statutory prompting, in order to secure commercial control by a celebrity over that individual's public persona (McKenna 2005). Technically, rights of publicity come out of the tort system, the legal regime for redress of personal wrongs and injuries, and so are not a form of property, but rather secure compensation to celebrities for commercial and dignitary injuries they suffer from unauthorized use of their public persona. As a practical matter, however, the right of publicity has evolved toward a full-fledged form of intellectual property, with the characteristics of exclusion, exchange, and disposition that come with a property regime. The right of publicity has been used by celebrities and by their estates to assert not only exclusivity over largely natural characteristics, such as facial image or tone of voice, but also over artificially constructed public depictions of the celebrity, such as the portrayal of Dracula by film actor Bela Lugosi, or the unauthorized use of distinctive catch-phrases associated with the celebrity to imply celebrity endorsement of a product.

In the *C.B.C. Distribution* opinion, the court held that the numeric representation of player performance was not sufficiently distinctive of their personality to ground a right of publicity claim. The court's holding focused on whether the use of the players' names in conjunction with their statistical records was the appropriation of a "symbol of their identity." In other words, the court considered whether the use of the statistics was representative of the individual, or was simply the use of the data as data.

The court concluded that use of such factual, historical data did not implicate the character, personality, reputation or physical appearance of the players.

In particular, the court distinguished previous similar cases as involving the appropriation of *likeness*. This analysis relied upon distinctions developed under the commercial right of privacy cases. The terminology of “privacy” is somewhat problematic; a “right of privacy” is in American law a term with so many aspects that it almost has no proper meaning. Various facets of the polyglot American right of privacy are drawn from scattered statutory or constitutional sources. Some of these deal with physical or bodily integrity, such as the constitutional rights against governmental invasion of the home or of reproductive activities. Others deal with personal data generated in the ordinary course of human activity: records of financial dealings, credit worthiness, social security identification, or medical history (Reidenberg 2000; Swire and Litan 1998).

In this particular instance, the court was concerned with a line of tort cases restricting the commercial use of a private individual’s likeness without that individual’s permission. These are the close cousin of the right of publicity cases, and early on the two lines of cases were closely intertwined, sharing similar precedent and reasoning. Such right of personal privacy claims were essentially the private citizen’s analog to the right of publicity claim; the one for the average person, the other for celebrities; one grounded in a right not to have one’s image commercially exploited, the other grounded in the right to control commercial exploitation that was routine and indeed expected as part of the celebrity’s career (Malkan 1997).

The court’s reliance on cases involving the common law right of privacy is critical because of the conceptual convergence between the private claim to prevent exploitation and the celebrity claim to control exploitation. The two lines of cases share a common lineage, but more importantly share a common set of rationales regarding the individual’s relationship to representations of likeness or identity. I have argued elsewhere that the law of privacy and the law of intellectual property comprise a matched pair of modalities for assigning control of data in modern information law (Burk 2006b). Here the convergence between the two is manifest in the analytical reliance upon privacy to resolve a question of publicity. While a distinction between the two kinds of cases is doctrinally correct, the distinction dissipates at a conceptual level, opening a window into the connected structures of information law.

The law of data representation

In certain respects, the *C.B.C. Distribution* suit was far from extraordinary. Protection of celebrity likenesses from unauthorized commercial appropriation has long been part of the American legal landscape, and celebrities have gone to court to prevent the appropriation of performance attributes ranging from distinctive singing styles to celebrated character roles to eponymous catch phrases. The right of publicity, the central legal theory in the case, has long been in a period of expansion, growing despite constitutional guarantees of free speech to cover the personas of deceased celebrities, parody or commentary on celebrity, and even to preclude allusions to celebrity (McKenna 2005).

But at the same time, this suit was in many respects quite extraordinary, in part because it may represent the high water mark for claims regarding celebrity attributes, a claim to control over a set of minimalist, publicly available statistics regarding player performance. The plaintiffs sought protection for numeric representations of player performance – in effect, for the most bare-bones kind of data about the players, for *facts* about the players. Not all data is factual of course, but the argument of the case equated the persona of the players with data constituting numerical representations of player performance. In seeking protection for facts about player performance, the plaintiffs appeared to push the right of publicity to its outer limits, and it is such limiting cases that reveal the underlying assumptions that structure intellectual property rights.

Under most branches of information law, facts would be considered unprotectable. For example, copyright is an unlikely regime to protect such personal data. Copyright extends only to works of original expression, that is, to creative expression originating from an author. The copyright statute excludes facts and other empirical representations, including many if not most databases, and most data representations in digital media can be considered as databases of varying complexity (Burk 2005). The creative arrangement of uncopyrightable elements can qualify for copyright, but this covers only the arrangement, not the arranged elements. And creative arrangements may be somewhat rare; the useful arrangement of data is frequently dictated by unoriginal considerations, such as alphabetical or numerical order, that will not qualify for copyright.

Despite the somewhat fanciful argument advanced by some commentators that the individual choices about the persona might constitute creative expression within the subject matter of copyright (Bartow 2000), personal facts are unlikely to constitute a

compilation for purposes of copyright. Even very complex distinctive features of an individual, such as facial features, are unlikely to be considered creative expression for purposes of copyright. The argument that bodily attributes are in some sense creative is not without foundation, as it could constitute a logical endpoint to some court decisions about the scope of copyright. I have observed elsewhere that certain recent copyright decisions have blurred the line between expressive choices and determined choices (Burk 2007). However, it seems unlikely that any court, even courts with an expansive view of copyrightable expression, would consider seriously either the choices made in an individual's life, or the attributes proceeding from such choices, as the type of original expression addressed by the copyright system. Such personal choices and attributes are far more likely to be classified as factual rather than expressive, and so excluded from copyright protection.

The law of trademarks lies closer in concept, if not in specific detail, to the type of legal structure that might recognize and accommodate interests in personal data representations. Trademark law incorporates into its doctrines a strong theme of protection the goodwill and reputation of businesses – that is, control over the public image and perception of a business and of the business' products (McKenna 2007). Thus trademark law is directed to indicia, such as words, phrases, logos, and other representations generated about a given business, in many ways paralleling a regime directed to the indicia and representations that might be generated about an individual. But this is only a parallel, not an exact match. Trademark law is specifically limited to the business context, and typically will not cover personal data. To qualify for trademark status, an indicator must designate the source of goods or services in commerce. Persona or personal identities are unlikely to qualify as goods or services, and are not necessarily used in commerce. Certainly in certain circumstances, such as that of celebrities whose persona is in many ways the commodity they sell, commercial information and personal information may merge, but the majority of data representations will fall into this category.

At some point such claims certainly intersect with, and perhaps conflict with, other regimes of control, such as the copyright or trademark systems. Actors such as Clayton Moore, Adam West, or Buster Crabbe may have been for many years associated with the fictional characters they portrayed in film or television: the Lone Ranger, Batman, or Tarzan. But those characters are also literary properties of the movie studios, the print media publishers, and

ultimately the authors who initially characterized those figures, and so subject to copyright of those creators. The characters may also be emblematic of the businesses that created and promoted products featuring those characters, and association of the characters with those businesses in the minds of the public may make the characters as trademarks of those businesses.

Thus it is no wonder that it was under the right of publicity, the commercial interest lying closest to the common law right of privacy, that the “fantasy sports” plaintiffs grounded their claim. The court nonetheless rejected the claim, much as it might have rejected a copyright claim to facts, due to the public nature of the statistics involved. But even though the court focused on the public nature of the data at issue, the more striking feature of the case may be the reductionist nature of the data at issue: the plaintiff's claims necessarily equate the identity of certain individuals with numerical data about those individuals, claiming that the data is sufficiently distinctive to constitute the persona of the players. The *C.B.C. Distribution* case thus provides a limiting case to consider where we might situate such factual representation in either information law or information ethics.

A bestiary of data representations

The issues attending data representations are in one sense nothing new. Player statistics long pre-date the current crop of information and communication technologies (ICTs), and the earliest American right of publicity cases stem from the unauthorized representation of sports figures on baseball trading cards that were packaged as a novelty with chewing gum. Portraiture, in marble or paint or celluloid film has a long history. Biographers back to ancient times depicted detailed textual portrayals of their subjects. Population census has been a fixture of social life as long as there have been taxes to collect and governments to collect them.

But what Floridi has dubbed the “radical re-ontologization” of the infosphere by ICTs has profoundly changed the nature and frequency of such representations (Floridi 2006). These examples of past data representations were largely limited to those wealthy or prominent enough to merit or afford them. The wealthy and famous are of course still more likely to be the subject of information records, but data representations are also now routinely generated and kept in a wide range of commonplace situations. In some instances, data representations are unintentionally generated by the everyday activity of the individual, or captured without the individual's

consent in a manner that might be objectionable were it noticed or fully drawn to the individual's attention. Records of credit card purchases or of surveillance cameras have become a routine feature of urban industrialized living. In other instances, development data representations are the product of acquiescence, as in the formation of medical records or the filling out of a credit application. The individual is cognizant and compliant, if not necessarily enthusiastic, about providing the personal data to be recorded.

In yet other instances, the creation of data representations are the product of active participation or directed activity on the part of the individual, whether posing for a portrait or developing a role-playing game avatar. Here it is likely that the individual is the most proactive in creating the representation of self. The degree of cognizance and intent in creating the representation seems bound up with the degree to which the data representation constitutes an information entity as Floridi has defined it. By contrast, unconscious or inadvertent generation of data, generated as the individual her activity is routinely monitored or recorded, seems least likely to reflect the autonomy integral to the definition of an information entity. These traces of informational activity often fall into the category that Westin (1968) refers to as the individual's "data shadow"; records which are generated by the passage of the individual through across a landscape of recording devices.

Floridi has relied upon Westin's distinction to separate data constitutive of the self from that which merely describes the activities of the self (Floridi 2005). But it is worth noting that this differentiation is problematic. Wiener famously described humans as "persistent information patterns" (1954, p. 96), but those patterns are part of and not separate from the larger infosphere; the human patterns bleed into the information environment while also taking their shape from it. It is not at all clear where to draw a line between constitutive and descriptive data; ostensibly descriptive data, about the individual or the individual's activities, might well be a major component of constitutive data, especially in the environment of ICTs. And the ability to assemble, process, and mine such records has begun to transform the data shadow into a full-fledged data portrait, in some cases revealing patterns of behavior and practice of which the subject may not be aware.

The development and increasing availability of digital media has greatly facilitated the creation, manipulation, and alteration of data representations – an aspect of the "radical re-ontologization" that Floridi contends ICTs have introduced into the information environment. This is true of representations at all levels of intentionality, but the effects on

the last category, the proactive formation of data representations, is of particular interest to this analysis. In some instances, digital media has made new data representations possible, as in the case of gaming avatars that allow the average person considerable latitude in developing and portraying a digital persona.

In many other instances, digitization and associated technological tools have allowed changes, sometimes drastic changes, in the character of data representations. For example, digital image processing famously allowed alteration of film footage from a classic Fred Astaire dance routine to insert new images, making it appear that the entertainer was dancing with a Dust Devil vacuum cleaner. Similar and increasingly more sophisticated techniques have been employed commercially, to make Gene Kelly, Audrey Hepburn, Humphrey Bogart and other deceased celebrities appear to endorse current consumer products, or artistically to place current performers such as Woody Allen, Tom Hanks or Jodie Foster into archival film footage so as to appear to interact with past celebrities. These techniques allow the fusion of two different representation technologies; traditional cinematic capture of an actor's character portrayal and computer image processing of that portrayal. The most recent generation of such techniques includes the digital sensor motion capture of an actor's character portrayal, which is then re-formulated to produce an entirely different appearance, in movies such as *The Lord of the Rings* or *Beowulf*.

The trend apparent in such visual representations, appears to be an increasing divorce of celebrity representation from association with a physical individual, accelerated by the capabilities of digital media. A similar progression of physically and graphically simulated portrayals can be seen in the trajectory of "virtual" performers in the music industry. The depiction of musical personas progresses from musicians specifically assembled into a band and assigned personas for the television show "The Monkees"; through the "Partridge Family," a group of television actors portraying a musical group, only two of whom performed musically, and whose music was largely dubbed by unseen and unrecognized backup musicians; to a group of unseen studio musicians providing music (including a hit single) for an animated television cartoon band the "Archies" to the elaborate invented personal histories and fictional personalities of the virtual band "Gorillaz." This progression approaches its apex in the digital creation of a Japanese "Virtual Idol" pop star DK96, known also as Kyoko Date. The logical endpoint of the trend is that described in William Gibson's science

fiction novel *Idoru* in which the pop star is an artificial intelligence, but that apex remains for the moment out of our technological reach (Gibson 1996).

The trajectory of such entertainment depictions suggests a difference of degree rather than one of kind; while each of these representations incorporates some aspect of an autonomous carbon-based intelligence, and is to some degree representative of that individual's personality or character, all involve a greater or lesser degree of calculated artifice. It is by no means clear that the personas created and depicted by physical actors in the "Monkees" are any more "real" or "genuine" than the personas drawn and depicted in the cartoon "Archies" characters. Certainly both were from a commercial standpoint successful extensions of the abilities of the individuals involved, and were undoubtedly valuable to their entertainment careers. To some degree they depicted some aspect of the individuals whose talents were incorporated into the depictions. But it is unclear to what extent either should be considered constitutive of the self, except perhaps that the cartoon seems more clearly separable from the underlying performer. In either case it is somewhat unclear what "I" might mean were either Davy Jones of the Monkees or Ron Dante the of the Archies to say "I had a hit single."

Each of these commercial personas would be encompassed to a greater or lesser degree by a network of intellectual property, including copyright and trademark protection, with the right of publicity likely becoming increasingly attenuated as the representations of identity become increasingly divorced from an actual person. Were someone to create an unauthorized depiction of Kyoko Date, any claim against the infringer would likely sound in copyright; until virtual performers are in fact AIs, and such AIs are recognized as personalities separate from their programmers, it seems unlikely that any court would recognize a right of publicity for such a performer. But it is also unclear if one were to mimic, say, the performance of the Archies, what right would be implicated – the right of publicity of the flesh and blood musicians who provided the sound for the cartoon band, or the copyright in the cartoon characters whose sound is provided by the musicians? Similarly it seems unclear what rights might be implicated by mimicking the of the Gollum character portrayed in Peter Jackson's cinematic rendering of *The Lord of the Rings* – leaving aside the contributions of the director and other cinematic collaborators, would an infringement implicate the rights of Andy Serkis, the actor whose movements were captured for digital manipulation into the Gollum image,

or those of the digital artists who transformed the data into an image altogether different than that of the actor?

A typology of data ownership

As the discussion to this point illustrates, information law must accommodate a wide range of data representations. We must sort them into some type of order before we can determine their amenability to analysis under information ethics. Fortunately, certain tools are readily at hand. The conjunction of intellectual property and privacy allows us to draw upon previous analyses of informational privacy in developing a more general theory of control over data representations. Particularly, in discussing the disposition of personal health and medical data, Annamarie Carusi (unpublished conference presentation, 2006) sets out a useful typology of data origination that bears upon the disposition of control over the resulting information. But this typology applies to the disposition of information generally, whether under the rubric of "privacy" or that intellectual property. The four approaches to data origination set out in the typology are useful in identifying, sorting through, and making sense of the regimes that might be applied to ownership and control of data representations.

The first approach in this typology adopts a posture of neutrality, assuming that data pre-exist in the universe, waiting to be found by human agents. Acts of measurement and recording are akin to capture rather than to creation. This approach assumes transparency for measuring and recording processes, that they simply reveal what is already "out there" to be discovered. This approach also assumes that data is not especially generated by or connected to a particular individual except insofar as the individual may be a phenomenon observed, like any other physical phenomenon. The implication of these assumptions is that data do not belong a priori to anyone; control may be assigned on the basis of capture or possession, but anyone would be free to capture or record the data from its natural state.

This approach is somewhat appealing due to its simplicity; its neutrality lends itself to a rule of "first possession" or perhaps "finders keepers." But the simplest rule may not be the most desirable rule, from the standpoint of fairness, or efficiency, or other distributive values we might prefer. Even if data are waiting to be found, certain personal information may have a social provenance that impacts the subject of that data. And, as we have seen, many types of personal data are actively created or fashioned by the

individual. The approach seems unduly naïve about the nature of measurement and recording, which since at least Heisenberg can hardly be viewed as separable from the influence of the observer.

Thus, a second approach might take into account the interests of the individual about whom data is collected. And, indeed, the law of privacy frequently views personal data as originating in the individual with whom the data is associated. Under this view, when an individual is tracked or assessed or measured, the data is effectively a product or an extension of that individual. The data belongs to the individual in the sense that it is associated with him; it comes from him and is in some sense a part of him. And of course it is only a short semantic move from saying the data belongs to the individual in an associative sense to saying that the data belongs to the individual in a propertized sense.

Alternatively, however, under a third view, the data generated about an individual might equally well be viewed as arising from the actions of the observer, created by the choice to observe or to measure. For example in the case of medical data, the data is generated by the tests and assessments performed on a patient by the physician. Without the actions of the physician, the data would not have been generated, and the character of the data is determined by the choices and judgments of the physician. The same might be said of other data, such as sports statistics; they could be considered the result of records compiled by fans or professional leagues or journalists. Patients and sports figures seldom record their own statistics; the vital signs of a patient or the efforts of a player would not be translated into data representations without the actions of an observer.

Viewing such data the result of the choices made by the observer shifts the associations underlying ownership and control of the data. When the focus shifts to the observer, our impulse is to assign ownership or control to the observer, who is now considered as the creator of the data. Certainly, there must be a patient or a player to be observed, but then there must be an object of observation in any measurement. If we view the data as the result of observation, there is no particular reason why we need assign control of the data to the observed individual, anymore than we would feel obliged to try to assign control of weather data to the atmosphere, or of cloud chamber data to the leptons and muons it records.

But this rationale cannot be entirely satisfactory. The observer and recorder of human data representations are not functioning in a moral vacuum. Unlike the observation of weather or subatomic particles, observation and measurement of another

human implicates the interests and autonomy of a sentient being other than the observer. Leaving the interests of the observed individual out of the calculus for control of data representation is just as problematic as ignoring the role of the observer in generating the data. Both an observer and an observed are necessary to the creation of the data; neither alone is sufficient.

Consequently, a fourth approach might be considered, an approach that views data generation systematically, as arising out of the interaction between the observer and the observed. Under this view, as the joint product of observer and observed, data reflects the choices made by each party to the interaction, and constitutes an extension of each. And, following the logic of assignment from the first two approaches, if ownership or control of data is to be given to the party that generated it, it follows that shared action implies shared control of the result. While this conclusion avoids the polar extremes of viewing either observer or observed in isolation, it entails its own practical problems. The application of this approach is not as clean or as clear cut as that of the other approaches. The logic of joint creation requires us to negotiate the tricky mechanics of structuring a regime for apportioning or sharing control over the information. While this approach offers a richer view of data origin, it may simultaneously result in a more complex assignment of rights, and so a greater propensity for uncertainty and for conflict over those rights.

The typology and the law

This typology of data origin is useful because each of these approaches has been seen not only within the law of informational privacy, but within the law of intellectual property. Thus, as I have mentioned above, copyright adopts the first approach in assuming that “facts” exist independently in the universe, and lie outside copyright because they are not the product of the author (Burk 2005, 2007). Some language in the *C.B.C. Distribution* opinion seems to adopt this approach for the right of publicity, suggesting that player statistics do not involve any protected aspects of the player persona, because they constitute independently available facts. But this argument simultaneously implies that where aspects of persona are involved, such as a player image, this stems from the individual depicted, and the right of publicity extends protection to that depiction: an adoption of the second approach.

The European approach to personal data privacy likewise adopts the second conception of data origin,

treating data as an extension of the individual or of the individual's personality. The EU directive on data privacy requires EU member states to enact legislation that essentially regards control over personal data as a fundamental right of the individual. The directive requires that individuals be notified about data collection, explicitly consent to the collection of their data, and grants them certain rights to review, correct, and purge data records (Reidenberg 2000). Data emanate from the individual, and so remain subject to the control of the individual, at least to a certain extent.

In contrast to this European approach, the American approach to personal data privacy effectively adopts the third conception of data origin. Consumer "data trails" of supermarket purchases, magazine subscriptions, telephone records, and credit ratings have generally been assumed to belong to the institutions, typically commercial institutions, that gathered or compiled the records (Swire and Litan 1998; Starke-Meyering et al. 2004). The recording entity has generally had broad discretion to buy, sell, trade, stockpile, and analyze such records, subject only to minimal constraints allowing the subjects of such records to "opt out" of certain types of trading or selling of their profiles. There are some sparse statutory restrictions on the use of records considered particularly sensitive, such as medical records or credit records. These few idiosyncratic statutes place somewhat more stringent controls upon dissemination of the records they cover. But these restrictions do not change the underlying assumption of American data records law: it is generally assumed that the records are properly in the custody of the recorder, almost completely under the control of the recorder, and constitute the property of the recorder. The individuals represented by such data have little or no entitlement to access or re-purpose the records. The role of the subject recorded is subordinated or unimportant; it is the action or investment of the recorder that is foregrounded.

Other data representations appear to fall within regimes that adopt the fourth approach. For example, I have previously argued that the proper ethical and legal framework for consideration of gaming avatars lies within copyright's regime for mixed origination (Burk 2006a). The ownership and control of gaming avatars, game items, and scenarios has been the subject of considerable discussion, and of occasional outright disputes between the players of role-playing games, and the providers of such games. For the most part, such disputes are controlled by contract law: before they will allow users to play their games, game owners typically require the player to assent to terms of use agreements that allocate copyright in any

characters or scenario features to the game owner. These terms grant the player a license to use the copyrighted works in the game, and may require an assignment of rights in return. Such agreements cannot override the definition of authorship under law of copyright – private agreements cannot change legislative enactments – but the contractual terms of use may re-assign whatever copyright claims the user may have to the game provider, as a condition of access.

Thus the user licenses for games provide a private answer to the question of ownership and control; an ownership contingency is transferred to the game provider before the game ever begins. But behind such licenses typically lies an assertion of copyright: one cannot determine whether a transfer is effective unless one knows what is available to transfer. Game providers have addressed the underlying assertion of copyright on the basis that the development of characters in the game scenario is attributable to the software options provided by the game owner. The providers argue that the avatars in a game, their actions and their attributes, are drawn from a suite of graphical depictions provided by the game owner. Avatars are stored as data, on the hardware of the game owner, and are translated into graphical and auditory depiction by the software of the game owner. The actions and storylines manipulated by players are reliant upon and derived from copyrighted works provided by the game owner: libraries of graphics and sound, sequences and ordered by the game software. Thus game providers argue that any character generated by game play lies within the expression of the game, within parameters specified by the game producer, and hence is subject to the game owner's copyright.

This logic is fine so far as it goes, but it cannot negate the contribution made by the player. Even if the choices for character development are finite, and supplied by the game owner, the number of choices is typically quite large, and the selection of character attributes and activities is almost certainly creative and original with the players, as such choices are viewed under copyright. The fact that the suite of choices is original with the game creator does not change the fact that the selection among the choices is almost certainly original with the players; both have had original input into the character that is formed from their contributions. This seems to place game avatars within one or more categories of shared authorship recognized by the copyright system. If the additions of original expression are authorized and sequential, then the result constitutes a derivative work of the game owner's initial work, and each contributor would hold copyright in her contribution

to the final product. If the contributors had an intent to produce a unified, integrated end product, then the result constitutes a work of joint authorship, and each contributor holds a complete and undivided interest in the resulting copyright. Depending upon the particular facts of a gaming situation, one or both of these designations could apply, but each of them attempts to recognize the interactive nature of the contributions from both game player and game owner.

The logic of the fourth approach seems also well illustrated by the view of the court in *C.B.C. Distribution* on player statistics that define fantasy sports teams. As I have already noted, these statistical notations are exceptionally sparse as data representations go. Taken in isolation, they have little or no meaning – they are simply a string of numbers that might represent almost anything: the measurements for shelving in someone’s hall closet; the gross national product of central African nations rounded to the nearest hundred thousand Euros; the numerical doodling of an idle mathematician. Without the addition of considerable contextual information, it is impossible to know which of these, if any, or of any myriad other possibilities the numbers might be properly be mapped. The correspondence of the data to measured phenomenon emerges only in the context of a newspaper sports page, or a sports almanac, or a fantasy sports competition. Only when juxtaposed against additional external information do the numbers begin to become meaningful as player statistical averages. The correspondence of the numbers to individuals is not inherent in the data; recognition of the numbers as somehow personal requires additional outside context to attach the data to particular sports figures.

It is worth noting that such decontextualization is routinely used in the law and practice of data privacy to provide sterile numbers for research or other manipulation. Concerns over autonomy and confidentiality are often considered alleviated through anonymization of the data, isolating it from particularized associations with individuals. Once such associations are gone, the numbers are no longer tied to a discrete identifiable individual, and so no longer considered subject to the consent or control of that individual. Such neutered indicia implicitly demonstrate the communal construction of personal data. Stripped of context, sensitive data is transformed from an extension of persona to free standing statistics that no longer implicate the individual with whom they were originally associated. The data are not essentially tied to the individual; such ties disappear with the context.

Such data represent the limiting case for the logic of the fourth approach. Different types of individual data will show this characteristic to a greater or lesser extent. Some data, such as sports statistics will be unrecognizable even as personal data unless labeled as such. Other data, such as some types of medical data, might be recognized out of context as individual data, but be unattributable to any given individual. Other data, such as facial images, may be immediately recognizable as individualized data belonging to an individual, even out of context, although locating the individual may be costly. These various types of data exist on a contextual continuum, but all show to some degree that in order for data to be meaningful data – the kind of data with which Floridi is concerned – it must be contextual data. It cannot stand alone. Meaningful data arises only in the context of interactivity; it is less meaningful, and perhaps even devoid of meaning as it becomes increasingly acontextual.

In a similar vein it is also with noting the overlapping, sometimes fragmented allocation of rights in data representations. At times, the assumptions adopted by different strands of information law may be at odds with one another. As mentioned previously, the common law right of privacy assigns to the private subject of a photograph a certain degree of control over the photograph’s commercial publication. Such rights align with the second of the three approaches outlined above. But at the same time, copyright in the photograph lies with the photographer, with the observer or recorder of the data. The assumption underlying copyright in photographs has been that the resulting image is constitutes the original expression of the photographer. Even if the image is of another individual, that individual is unlikely to be regarded as the “author” of the image; rather, it is the photographer’s creative selection of lighting, camera angle, exposure, and so on that is reflected in the assignment of ownership of the image. The exclusive rights granted by copyright may be modulated or overridden by overlapping rights granted to the subject of a photo: rights of publicity or of privacy may sometimes give the subject a veto over certain uses of the image. Such overlapping or incomplete rights effectively create a regime of shared control. The practical effect of these overlapping rights runs contrary to the paradigm of property law, which for both real and intellectual property has assumed the “sole and despotic” control regime of real property as its archetype (Burk 2003). But for purposes of copyright, the third approach, holding that the data representation properly belongs to the observer, is the approach that prevails.

The ethics of information law

In previous work I have noted that both privacy law and intellectual property law display one of two dominant rationales, characteristic of the jurisdiction in which they originate (Burk 2006b). In the modern global environment, the dominant information law paradigms arise in either the United States or the European Union, the former grounding its information law in a largely utilitarian rationale that focuses upon the economic benefits of information policy, the later grounding its information law in a larger deontological rationale that emphasizes the autonomy of the individual. These two ethical frameworks in large measure define the terms for modern information law.

It is critical to recognize that although these ethical values reflect the common structures of information law, they do not fit neatly into the typology already outlined above. It would be appealing to assume that because the second approach of the typology focuses upon the individual observed, that it tends to uphold the value of autonomy; while because the third tends to focus on the observer, it tends to focus on incentives that uphold more utilitarian values. But blithely mapping these structural frameworks to an accompanying ethical framework would be a mistake. Both utility and autonomy are two edged swords, and either the second or the third approaches – and for that matter, possibly the first and fourth – might be pressed into service to promote either value.

In general, from a utilitarian perspective, incentives to generate valuable data should be placed with the party most likely to respond to the incentives, granted to the party most likely to invest in the creation of valuable data when offered the reward of ownership. It is not entirely clear how this principle will play out in the case of an individual's information. First, as indicated above, a good deal of information is recorded or thrown off unconsciously or involuntarily, as a routine activity or a "data shadow." It is not at all clear that we need to encourage additional investment in such activity, or that we would succeed if we tried. Other data representations are the result of conscious activity on the part of the individual.

It may be that there is value to be gained by encouraging not the generation of personal data, but the assembly of such data into collections. But here it becomes clear that an individual's data is a peculiar commodity, often subject to market failure. For the most part, personal data is a valuable market commodity only in the aggregate. (Stark-Meyering, Burk, and Gurak 2004) Sports statistics aside, typically no individual datum is worth very much; even the

collection of data associated with a given individual is worth very little. The value of such individual data is likely far less than the cost of negotiating a price for it; it likely costs more to sell the data than the data is worth, so no market for individual data is likely to form.

Even in the case of celebrity data, where images and persona are worth bargaining over, it is unclear where a utilitarian calculus will lead. Statistics about player performance are valuable to the athletes, to their teams and their leagues. But the data are also clearly valuable to the fans, who enjoy comparing, discussing, and yes, playing fantasy sports with the data. The number of fans is far larger than that of the players, and a little value on the part of each fan may in the aggregate exceed the value placed upon the data by the player it depicts. Indeed, the fans may suffer from a collective action problem: they may value the data more as a group, although individually less than the professional leagues, and if the fans have no mechanism to express the group valuation, it may go unappreciated.

Adopting autonomy as a goal is leads to equally inchoate results. From the standpoint of autonomy, ownership or control surely belongs with the individual whose personality is invested in the data. But this logic begs the question of which individual that might be. Within existing information law, for example, the law of privacy and the law of copyright point in different directions; the former assumes that it is the observed individual who is most invested in the data, so that respect for autonomy is best served by deferring to that individual's preferences. But the law of copyright assumes that the observer, the photographer or painter or biographer, is the author who is invested in the resulting work, and that autonomy is best served by deferring to that individual's preferences. The law of privacy and publicity tell us on the one hand that we cannot use the observed individual as a means to effectuate the observer's purposes, but copyright tells us on the other hand that we cannot allow the observer to be used as means to effectuate the observed individual's ends, either.

As a practical matter, the two bodies of law may well be able to point in different directions. There need not be direct doctrinal conflict between the two regimes; copyright does not extend to unoriginal facts about the observed, while privacy law may. But the assumptions underlying the assignment of ownership or control in each instance are not so easily reconciled when extrapolated to new and ubiquitous forms of personal representations, as for example when the holder of detailed data profiles wishes to sell the profiles for targeted marketing. We have seen that we

likely will receive little concrete guidance on the allocation of control over such profiles from the existing ethics of information law. The question remains whether information ethics gives a clearer guide to structuring allocations of control over data representations.

Applying information ethics

Having surveyed the approaches adopted under various branches of information law, we are in a position to consider how the doctrine in these areas might align with a theory of information ethics. The outlook is fairly mixed. Areas of information law, such as European data privacy, or the right of publicity, that have adopted the second model would appear to lend themselves readily to Floridi's information ethics, as they tend to assume that the data representations they address arise from the individual and are constitutive of the individual. But accepting such compatibility too quickly would be a mistake. As I have already indicated above, there remains in the theory of information ethics a question as to what data constitutes the individual, and what is simply data about an individual's actions. Given this distinction, we might wonder whether these apparently compatible doctrines *should* assume that the information they contemplate is constitutive. For example, it may be that some of the information assigned to the individual under data privacy laws is merely part of the "data shadow" that is not constitutive data, and so might just as properly be assigned to the control of others.

By the same token, we are left in a quandary when testing the converse proposition. Information ethics as presently constituted offers little guidance to test whether there is a proper fit with areas of information law that do *not* assign control to the individual, in determining whether the areas that have adopted the first, third, or fourth models address data that is *not* constitutive of the individual, and so permissibly assign control elsewhere. Certainly in many cases the data representations that are left to the control of someone other than the individual represented seem highly individual, unique, and identifiable. But are these the characteristics of constitutive data? Statistics about sports figures provide perhaps the most sparse data representations possible, yet they are distinctive, and information ethics does not seem to assist us in determining whether the court in *C.B.C. Distribution* correctly excluded them from the right of publicity or not.

Thus, at least to date, the theory of information ethics leaves open this central and exceptionally

difficult question as to what data is or ought to be considered the individual's data. To say as Floridi has that invasion of privacy is akin to kidnapping or assault rather than to trespass illustrates the point that information is personal, rather than separate from the individual. But it does not do so at a very deep level. Much of trespass is about personality even in the world of atoms; as Radin points out, stealing a wedding ring is a conversion or trespass, but this is not necessarily because of the market value of the ring, it is because ring is an object by which the owner has defined himself and is constitutive of his personality (Radin 1982). The quick tour of data representations I have made here shows this to be even more common in the world of bits; data representations are very often informational objects by which individuals define themselves. Copyright and the right of publicity may protect against forms of informational trespass, but they are also very often very much about protecting the identity of the individual (Heyman 2005).

To be sure, Floridi recognizes that information may "overlap," noting that if John is married to Mary, this is information about both John and Mary (Floridi 2005). But this treats the information as something in the air, something hanging independently in space. We have seen that the law sometimes adopts this stance, but treating the information as *res nullius* still requires some theory as to who may appropriate it, or where it is to be assigned. Just as often, law treats information as constitutive of the personality to whom it is attached; this seems more in keeping with the constitutive theory of information entities Floridi explicitly espouses. But even this stance requires some theory as to which information is constitutive of the individual. The information captured in a portrait is as much about the artist that painted it as it is about the individual who sat for it, and law requires some theory of assignment when conflicts between the interests of the two arise.

This gap in application poses a particular problem with regard to the fourth model, where the data is meaningful only in context, or where the generation of personal data representations is a shared endeavor. Such data representations are not merely data that overlaps individuals, but data that would not exist without the input or contribution of multiple individuals. And this may prove to comprise the majority of cases; indeed one might argue that there is no data representation that it not the result of shared or communal interaction. Copyright law partially recognizes this in the concept of joint authorship; other branches of information law assign fragmented control of data representations, to one party in one case, to another party in another.

A coherent theory of information ethics might be enormously helpful in sorting out this tangle: whether assignment of data should be split, and if so when, and how. But this requires a more robust model of what information constitutes the individual entity's information, and that crucial aspect of the theory remains undeveloped. So long as this is the case, the theory will remain one of very limited legal application.

Conclusion

I have shown in this essay that in several respects the law addressing rights of privacy, rights of publicity, and related intellectual property rights already treats or tends toward treating personal data representations as constitutive of individuals, with the potential of treating them as information entities. It would be a relatively small adjustment in such cases to view the representations covered by such law as Floridian information objects, amenable to his theories of information ethics. But in other respects the law of privacy and intellectual property is ethically incoherent, or points in a different direction, adopting disparate models to assign responsibility and control for data representations. A theory of information ethics might lend coherence to these bodies of law, harmonizing the treatment of data across a variety of legal doctrines. As matters stand now many aspects of information law would require re-alignment in order to be compatible with such a theory.

This lack of coherence of course stems in part from general incongruities in the law itself. Development of legal doctrine occurs unevenly as a matter of necessity; as Justice Holmes famously remarked, the life of the law is not logic, but experience (Haack 2007). But the lacunae in such an approach to information law are not merely due to legal inconsistencies: it is by no means certain that information ethics itself is yet up to the task or providing a global theory for information law. In particular, it is critical to determine under Floridi's framework what information constitutes the self and what does not. The constitutive claim somewhat begs the question. To say that an information entity is constituted by his information assumes that we know what information is "his" or "her" information; it assumes at least a theory of association if not of ownership.

Although Floridi has begun to address certain of the challenges that a robust theory of information ethics will face (Floridi 2006), this initial "test drive" of information ethics through the legal landscape reveals some additional challenges that become apparent at the point where information ethics is applied to legal relationships. A richer theory of

information ethics promises a coherent framework in which to align the disparate branches of information law, but it can only do so if it offers some guidance as to whom information belongs, either constitutively or proprietarily. Thus application of information ethics to information law must await a fuller theory of information ethics, but perhaps recognizing the current gaps will assist the development of a fuller theory.

References

- A. Bartow. Our Data, Ourselves: Privacy, Propertization and Gender. *University of San Francisco Law Review*, 34: 633–704, 2000.
- D. Burk. Legal Consequences of the Cyberspatial Metaphor. In M. Consalvo et al. editors, *Internet Research Annual Vol. 1: Selected Papers from The Association of Internet Researchers Conferences 2000–2002*, pp. 17–24. Peter Lang, 2003.
- D. Burk. Expression, Selection, Abstraction: Copyright's Golden Braid. *Syracuse Law Review*, 55: 593–618, 2005.
- D. Burk. *Electronic Gaming and the Ethics of Information Ownership 4*. International Review of Information Ethics 4, 2006a. <http://www.i-r-i-e.net/inhalt/004/burk.pdf>.
- D. Burk. Privacy and Property in the Global Datasphere. In: S. Hongladarom and C. Ess, editors, *Information Technology Ethics: Cultural Perspectives*, pp. 94–107, Idea Group, 2006b.
- D. Burk. Method and Madness in Copyright Law. *Utah Law Review*, 2007: 587–618, 2007.
- C.B.C. Distribution and Marketing, Inc. v. Major League Baseball Mass Media, L.P., 443 F. Supp. 2d 1077 (E.D. Mo. 2006).
- L. Floridi. Information Ethics: On the Philosophical Foundations of Computer Ethics. *Ethics and Information Technology*, 1: 37–56, 1999.
- L. Floridi. On the Intrinsic Value of Information Objects and the Infosphere. *Ethics and Information Technology*, 4: 287–304, 2003.
- L. Floridi. The Ontological Interpretation of Informational Privacy. *Ethics and Information Technology*, 7: 185–200, 2005.
- L. Floridi. Four Challenges for a Theory of Information Privacy. *Ethics and Information Technology*, 8: 109–119, 2006.
- L. Floridi and J.W. Sanders. Artificial Evil and the Foundation of Computer Ethics. *Ethics and Information Technology*, 3: 55–66, 2001.
- W. Gibson. *Idoru*. Penguin, 1996.
- S. Haack. On Logic in the Law: "Something but not All". *Ratio Juris*, 20: 1–31, 2007.
- L. Heyman. The Birth of the Authornym: Authorship, Pseudonymity, and Trademark Law 80. *Notre Dame Law Review*, 80: 377, 2005.
- J. Malkan. Stolen Photographs: Personality, Publicity, and Privacy. *Texas Law Review*, 75: 779–835, 1997.

- M. McKenna. The Right of Publicity and Autonomous Self-Definition. *University of Pittsburgh Law Review*, 67: 225–294, 2005.
- M. McKenna. The Normative Foundations of Trademark Law. *Notre Dame Law Review*, 82: 1839–1916, 2007.
- N. Negroponte, *Being Digital*. Vintage Books, New York, 1995.
- M.J. Radin. Property and Personhood. *Stanford Law Review*, 34: 97–1015, 1982.
- J. Reidenberg. Resolving Conflicting International Data Privacy Rules in Cyberspace. *Stanford Law Review*, 52: 1315–1371, 2000.
- D. Starke-Meyering, D. Burk, and L. Gurak. American Internet Users and Privacy: A Safe Harbor of Their Own? In P.E.N. Howard and S. Jones, editors, *Society Online: The Internet in Context*, pp. 275–294. Sage, 2004.
- P. Swire and R. Litan, *None of Your Business: World Data Flows, Electronic Commerce, and the European Privacy Directive*. Brookings, Washington DC, 1998.
- N. Wiener. *The Human Use of Human Beings: Cybernetics and Society*, (2nd ed.). Doubleday Anchor, 1954.
- A.F. Westin, *Privacy and Freedom*. Atheneum, New York, 1968.