JOURNAL OF LEARNING ANALYTICS
SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH
(2014). Contemporary Privacy Theory Contributions to Learning Analytics. *Journal of Learning Analytics*, *1* (1), 140–149.

Contemporary Privacy Theory Contributions to Learning Analytics

Jennifer Heath
University of Wollongong
Australia
jheath@uow.edu.au

ABSTRACT: With the continued adoption of learning analytics in higher education institutions, vast volumes of data are generated and "big data" related issues, including privacy, emerge. Privacy is an ill-defined concept and subject to various interpretations and perspectives, including those of philosophers, lawyers, and information systems specialists. This paper provides an overview of privacy and considers the potential contribution contemporary privacy theories can make to learning analytics. Conclusions reflect on the suitability of these theories towards the advancement of learning analytics and future research considers the importance of hearing the student voice in this space.

KEYWORDS: Learning analytics, privacy theory, privacy

1 INTRODUCTION

The anticipated benefits of learning analytics in higher education are well documented and frequently focus on data issues and technical matters associated with system development and implementation. As Willis, Campbell, and Pistilli recently discussed:

> Big data and analytics, which marries large data sets, statistical techniques and predictive modelling [to mine] institutional data to produce 'actionable intelligence' present big questions to those of us in higher education. (2013)

This paper focuses on the privacy aspects of learning analytics deployment as a component of the ethical dimension of learning analytics. This paper is written from the position that having the technical capability to conduct a particular learning analytics task does not automatically mean that the task should be performed. As the discussion here will outline, there are many facets to privacy and some of the older concepts may not adequately serve the needs of learning analytics stakeholders, including academics, institutions, technology providers, and — most importantly — students.

2 PRIVACY, WITH A FOCUS ON INFORMATION PRIVACY

Philosopher Herman Tavani provides an insightful phrase that is a useful starting point for considering privacy matters: "Privacy is a concept that is neither clearly understood nor easily defined" (Tavani, 1999, p. 11). Publications concerning privacy matters in Western culture have been provided across multiple disciplines, including law and philosophy (Cohen, 2000; Fried, 1968; Rachels, 1975; Warren & Brandeis, 1890; Westin, 1967) and those with a focus on information

privacy (Floridi, 2005, 2006; Kang, 1998; J. Moor, 2000, 2005; J.H. Moor, 1997; Nissenbaum, 2010; Shoemaker, 2009; H. Tavani, 2007; Tavani & Moor, 2001; H.T. Tavani, 2007). Figure 1 presents a very simplified overview of privacy in order to provide some insight into the foundation concepts of privacy across four broad areas.

The first area uses a concise conceptualization of privacy from Culver et al. (Culver, Moor, Duerfeldt, Kapp, & Sullivan, 1994) where they argue that a person can be said to have privacy if, in a given situation or context, he or she is offered protection from *intrusion*, *interference,* and *information access* by others. This conceptualization of privacy is similar to that raised by Warren and Brandeis (1890) in their seminal paper on the rights of an individual to be left alone and free from intrusion and interference. The second area describes two broad classifications that also assist in the conceptualization of privacy: being normative and descriptive privacy. In a normatively private situation, individuals are protected by cultural norms such as formal laws or informal policies. Normatively private situations often include zones or contexts where normative protection is needed, for example, a patient in consultation with a clinician or a client in discussions with a lawyer. Descriptive privacy results in situations where individuals can expect privacy by natural means such as physical barriers. There are additional suggestions of dichotomies of "personal" versus "public" and zones of privacy (Gerstein, 1984) with privacy expectations varying according to the classification of the information type.

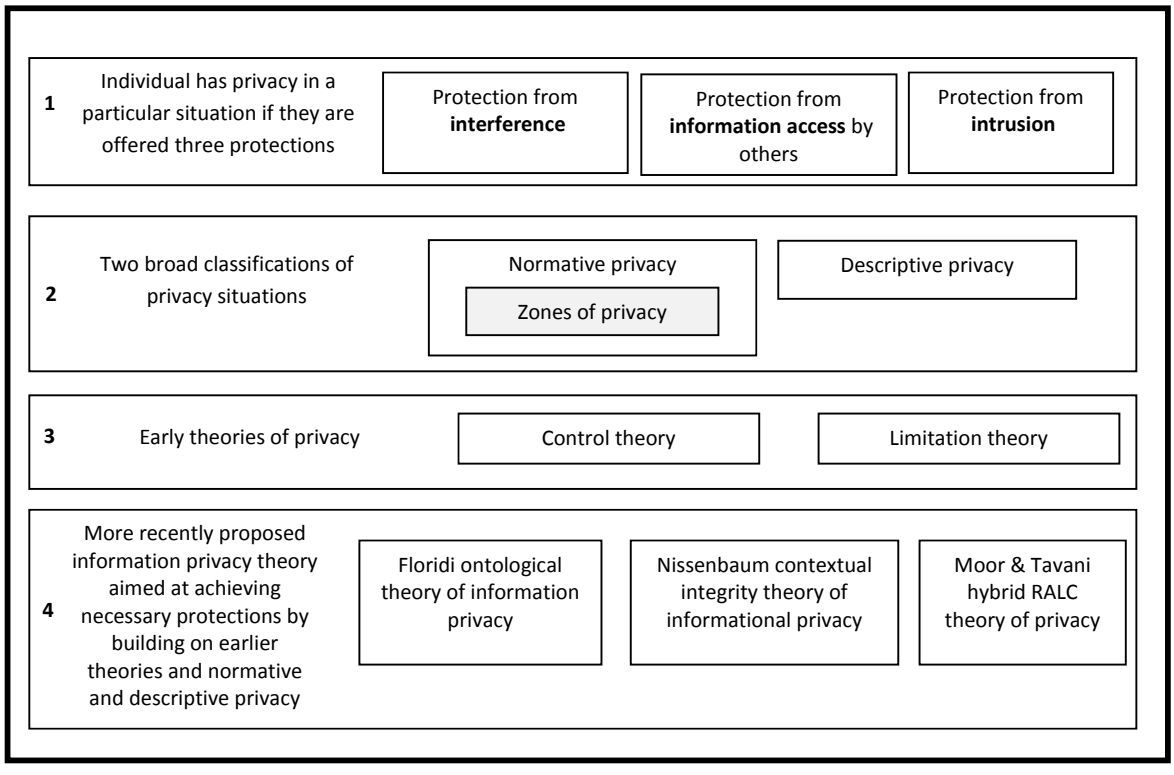

Figure 1: Broad overview of privacy

The third area, in Figure 1, presents two early theories of privacy that focus on allowing individuals control over their personal information, or limitations on the persons who could gain access to personal information. Debate regarding privacy has swung between arguments for and against a particular approach with the limitation theory and control theory dominating. Some publications in the learning analytics domain refer to privacy matters and take a single perspective from Figure 1 as the guide. For example, Petersen and Worona use a control theory interpretation of privacy to suggest how privacy affects learning analytics: "Privacy relates to the ability of individuals to control information about themselves" (2006, p. 16).

The final area includes three contemporary theories of informational privacy that move beyond the control theory versus limitation theory debate and offer a more holistic approach to privacy where the context (or *infosphere* in Floridi's work) emerges as a very important component of privacy theory. Luciano Floridi (2005) proposed an ontological theory of informational privacy based on information ethics. In a follow-up publication, Floridi (2006) provides a concise summary of his theory:

> To summarise: given a certain amount of personal information available in (a region of) the infosphere *I*, the lower the ontological friction in *I*, the higher the accessibility of personal information about the agents embedded in *I*, the smaller the informational gap among them, and the lower the level of informational privacy implementable about each of them. Put simply, informational privacy is a function of the ontological friction in the infosphere. (Floridi, 2006, p. 110)

Applying Floridi's privacy theory to the real world would be possible; however, the more tangible, less esoteric nature of Nissenbaum's, Tavani's, and Moor's theories provide a useful bridge to the "real world" of learning analytics.

3 CONTEMPORARY PRIVACY THEORIES AND APPLICATION TO LEARNING ANALYTICS

Contemporary privacy theories proposed by Nissenbaum (2010), Tavani (H. Tavani, 2007; Tavani & Moor, 2001; H.T. Tavani, 2007) and Moor (J. Moor, 2005; J.H. Moor, 1997) have been developed with the intention of applying them to diverse contexts, such as the rich learning analytics environment.

3.1 Nissenbaum

Nissenbaum moves the privacy debate beyond "control" or "limitation" theory, stating:

> Common usage suggests that intuitions behind both the constraint and control conceptions are sound: namely, that control over information about oneself is an important dimension of privacy, but so is the degree of access that others have to this information, irrespective of who is in control... In my view, the effect of these challenges, coupled with persuasive arguments, is not to prove that one or the other of these approaches is correct, but that both capture essential aspects of

> privacy that we seem to care about. A non-arbitrary resolution of this disagreement is not possible. (Nissenbaum, 2010, p. 71)

Nissenbaum proposes "contextual integrity" as an alternative conception of information privacy. Her approach is comprehensive with a goal of providing a decision heuristic to guide the evaluation of information privacy, which has potential benefits for learning analytics environments.

> … a right to privacy is neither a right to secrecy nor a right to control but a right to appropriate flow of personal information … Privacy may still be posited as an important human right or value worth protecting through law and other means, but what this amounts to is contextual integrity and what this amounts to varies from context to context. (Nissenbaum, 2010, p. 127)

Nissenbaum proposes informational norms that govern activities in contexts that she refers to as "context-relative informational norms." These norms are characterized by four key parameters: (1) contexts, (2) actors, (3) attributes, and (4) transmission principles. Nissenbaum provides a comprehensive definition of contexts as "structured social settings characterized by canonical activities, roles, relationships, power structures, norms (or rules), and internal values (goals, ends, purposes)" (2010, p. 132). In the learning analytics context, the broader higher education environment canonical activities, roles, relationships, and so on are apparent. The learning analytics environment is not a static context. One context in which the student may engage is with the learning management system (LMS) where exchanges take place between academic staff and students engaged in learning. Frequently student involvement is mandatory in this context. The internal values (goals, ends, purposes) of student engagement with the LMS relate to providing a stimulating learning environment and effective management of student engagement. As students engage with online activities, data is generated as a by-product of this activity, including patterns of questions posed and answered (Buckingham Shum & Ferguson, 2012). This data does not inform measures related to learning outcomes as collected by assessment items, but does provide valuable insight into student learning engagement. Where does this sit with the internal values of the LMS context?

Another relevant context is associated with application, admission, and administration of the student journey into and through the Institution. Again, this context sees students required to provide personal information in order to progress through their administrative matters. The internal values (goals, ends, purposes) in this context tend to focus on the efficient management of student administration matters.

The above two contexts operate with different internal values, however learning analytics frequently merges the data from the two contexts into consolidated datasets ready for analysis where yet another set of internal values are encountered, the precise nature of which is emerging and under discussion at many Institutions. The definition of learning analytics provided by Ferguson, as reported by Ellis (2013), is a reasonable indicator of the goals, ends, and purposes of learning analytics: "the measurement, collection, analysis and reporting of data about learners and their contexts, for the purposes of understanding and optimising, learning and the environment in which it occurs." Given the mandatory requirement for students to engage in the above two contexts, is it

adequate to assume that they provide tacit agreement with the goals, ends, and purposes of this new context?

Second, Nissenbaum identifies three types of actors: (1) senders of information (2) recipients of information, and (3) information subjects. Table 1 provides an overview of key actors in the learning analytics domain. An "X" indicates typical higher education stakeholders and the actor roles they adopt in learning analytics contexts.

Table 1: *Privacy in Context*: Actors (Learning Analytics environment)

Actors	Individual students	Collaborative groups of students	Academic staff – subject coordinators	Academic staff – tutors, facilitators etc	Information Technology professionals	University Administrators, business analyst, planners
Senders of information	X	X	X	X	X	
Recipients of information	X	X	X	X	X	X
Information subjects	X	X	X	X		

It is interesting to note that academic staff may also be considered information subjects as the details of their engagement with students, including assessment marking and comments, may be information of interest to other actors. Diaz et al. (Diaz, Golas, & Gautsch, 2010) suggest that academics are more concerned about privacy than students are. The surveillance dimensions related to academic staff are interesting and require further research.

About attributes (information types) Nissenbaum says, "Analysis of attributes in contextual integrity is more nuanced than the private/public dichotomy of information. Informational norms render certain attributes appropriate or inappropriate under certain conditions and attributes co-evolve with contexts" (p. 132). In the learning analytics domain, attributes are diverse and constantly evolving, hence Nissenbaum's recognition that they co-evolve with contexts is an important aspect of her privacy theory. This pairing re-enforces the dynamic nature of the always-evolving environment and the need for privacy theory to keep pace. A static, rigid approach to privacy is inadequate in the learning analytics (and many other) technology-enabled activities.

The idea that privacy implies a limitation of access by others is similar to Nissenbaum's concept of an informational norm. In Nissenbaum's theory, diminishment of access is just one way that information flow may be governed. She defines transmission principles as "a constraint on the flow of information from party to party in a context" and the "terms and conditions under which such transfers should occur" (p. 132). In the two contexts described above, student administration and learning-management system context, the students enter into a tacit agreement regarding the transmission principles. For example, by submitting an assignment in a learning management system, students are allowing the information to flow from themselves to the academic staff member who will provide feedback on the assessment item. The flow of information is from the student to the responsible academic and back again. The terms and conditions under which this

information flows is an information norm when engaged in a learning context facilitated by a learning management system.

The transmission principles related to the flow of information from individual students to information technology professionals or University administrators is not necessarily an information norm. Terms and conditions under which these transfers should occur begin to assist in unpacking the complex ethics and privacy issues surrounding learning analytics.

Transmission principles regarding the provision of a student's personal demographic data in the student application, admission, and administration context do not necessarily apply in any other context. If a student agrees to the flow of student equity-related data to support admission processes, he or she is not necessarily agreeing to the same terms and conditions of information flow in another context, such as secondary use of data for learning analytics activities.

The two following scenarios, using the above-described four key parameters, illustrate the variations that can occur across learning analytics initiatives and hence the variation in the breadth and depth of privacy matters:

> **Scenario #1**: Analytics visualization for student use. Colour-coded indicators displayed to individual students to provide clear visualization of their progress in completion of assessments within a subject.
>
> **Context**: Provision of personalized information to individual students during semester using the academic data for specific subjects typically available in learning management systems.
>
> **Actors:** Sender of Information is the teaching academic, Recipient of Information is the individual student, and the Subject of Information is the individual student.
>
> **Attributes:** Assessment data (e.g., name, due date, learning outcomes), Student assessment progress data (e.g., date submitted, assessment mark, days overdue, learning outcome achievement).
>
> **Transmission Principles**: Data flow terms and conditions: Tacit within the teaching-learning relationship that exists between academic and student. No change in context affects this scenario as the data is generated and used within the one context, which is the learning within a particular subject instance.
>
> **Scenario #2**: "At risk" student modelling and associated interventions. Informed by predictive analytics modelling that includes diverse datasets from multiple university transaction processing systems, including student demographics, admission pathway, engagement with support services (including student well-being and academic type support services), attendance records from labs, tutorials, and myriad of data captured when student engages with the learning management system.

Context: Broad use of transaction processing information generated as student engages with mandatory and optional administrative and support services across the university environment.

Actors: Senders of Information are the custodians of diverse information systems, Recipients of Information are the academic or professional staff responsible for using "at risk" predictive models and initiating the interventions for individual students; Subjects of Information are the individual students.

Attributes: The data comprises all the "electronic breadcrumbs" left by students as their higher education journey moves from application to admissions, enrolments, and engagement across the institution.

Transmission Principles: Data flow terms and conditions from the original context where information systems gathered student data — say at the application stage where potential students provide demographic data — are quite different from the data flow terms and conditions that must be considered in the context involving building "at risk" models and encouraging staff to intervene with students. As the data subjects, the students should, ideally, have influence over the data flow terms and conditions, including options to remove themselves from this modelling and intervention scenario.

The descriptions above are a first step in unpacking the potential for Nissenbaum's contemporary privacy theory to assist in the analysis of privacy scenarios in the learning analytics domain.

3.2 Tavani and Moor

The work of Tavani and Moor also assists in navigating through this grey area of privacy and learning analytics. Through a series of individual and jointly authored publications, Tavani (Tavani, 1999; H. Tavani, 2007; Tavani & Moor, 2001; H.T. Tavani, 2007) and Moor (J. Moor, 2000, 2005; J.H. Moor, 1997) proposed a hybrid privacy theory that seeks to move beyond early privacy theories. The result is an identification of the fundamental, essential components necessary in a privacy theory. One outcome of their research is a tripartite model to describe a sufficient theory of privacy that they suggest must include three core aspects: (1) concept of privacy, (2) justification of privacy, and (3) management of privacy:

> A good theory of privacy has at least three components: an account of the concept of privacy, an account of the justification for privacy, and an account of the management of privacy. This tripartite structure of the theory of privacy is important to keep in mind because each part of the theory performs a different function. To give an account of one of the parts is not to give an account of the others. (Tavani & Moor, 2001, p. 6)

Moor and Tavani tackled the fundamental, important matter of developing a privacy theory rather than devising particular justifications or recommendations for the management of privacy suitable

for particular contexts. This approach thus creates a privacy theory foundation that will keep pace with technological innovations and supports the future directions of learning analytics activities. The resulting theory can be effective in a wide range of contexts with sufficient provision to respond to constantly developing technologies that could bring insufficient conceptualizations of privacy undone.

In a practical scenario, the "concept" and "justification" of privacy related to learning analytics can be addressed through careful consideration and clear articulation in learning analytics governance policy. An institution may choose to adopt a particular philosophy to underpin consideration of these privacy matters. The concept of privacy in the learning analytics domain broadly encompasses protection from intrusion and data gathering by actors who are not the subject of the information (i.e., individual students). The justifications of privacy are often "rights" based with the rights of both students and academic teaching staff to be considered within the learning analytics domain. The "concept" and "justification" of privacy are reasonably stable in the higher education learning analytics domain. Each of these aspects of privacy should be addressed in the learning analytics governance policies of higher education institutions, ideally before adopting learning analytics strategies. The scope of "management" of privacy in learning analytics scenarios includes the combination of technologies, policies, and procedures designed to address learning-analytics privacy requirements. With the emergence of new technologies, the "management" aspect of learning analytics privacy will be more volatile than the comparatively stable "concept" and "justification" aspects.

4 FUTURE CHALLENGES

Another recent publication proposes six principles for an ethical framework for learning analytics (Slade & Prinsloo, 2013) and three of these principles intersect with contemporary privacy theory, specifically the following: P2, Students as agents; P3, Student identity and performance are temporal dynamic constructs; and P5, Transparency. *P2, Students as agents,* encourages the view that students are collaborators in learning analytics and should be involved in decisions regarding use of their data. This is similar to the recognition of Actors (Information subjects) and Transmission Principles (Nissenbaum, 2010) and management of privacy expressed through choice, consent, and correction (Tavani & Moor, 2001). *P3, Student identity and performance are temporal dynamic constructs,* and *P5, Transparency,* speaks to the management aspects of privacy. Future challenges surround the effective integration of proposed ethical frameworks with valuable theoretical foundations, such as those proposed in contemporary privacy theories.

The rapidly evolving field of student co-creation of material (Diaz et al., 2010) presents fresh challenges as learning materials shift from being distributed by static online tools to the involvement of third-party providers (Rotenberg & Barnes, 2013). For example, how do third-party providers sit with the actors depicted in Table 1? What transmission principles guide the flow of data to third-party providers? The outcomes of recent research (Drachsler & Greller, 2012) indicate stakeholder concerns regarding intellectual property. How, then, does this rest with co-creation of materials? Drachsler and Greller (2012) described privacy as a "soft barrier" to learning analytics and investigated the opinion of education practitioners and researchers (n=156) in relation to privacy. It

is not clear if a shared definition of privacy was provided to participants and, as has been highlighted in this paper, there are diverse perspectives and interpretations of privacy. The four questions covered are as follows: breaches of privacy and intrusion in personal affairs; ethical principles around sex, political and religious beliefs, and ethnic origin; ownership rights and intellectual property (IP); and freedom of expression. Focussing on the first question, 65.8% of respondents indicated an expectation that learning analytics would affect privacy and personal affairs. Results from question three indicate that 60.1% of respondents indicate that ownership and IP would be affected by learning analytics. Respondent opinion was less clear regarding questions two and four.

This "soft barrier" study did not engage with students but focussed on the opinions of educators and researchers. The contemporary privacy theories considered here clearly recognize the importance of data subjects in determining appropriate privacy solutions. In the learning analytics domain, students are, as illustrated in Table 1, important actors and their voices need to be heard. Therefore, future research must engage with students in order to hear their expectations and concerns about privacy matters regarding advancing learning analytics.

5 CONCLUSION

Contemporary privacy theories can make a valuable contribution to learning analytics by providing clearly articulated, comprehensive conceptualizations of privacy. The theories explored here provide guideposts for considering the privacy dimensions of scenarios from the visualization of assessment progress data by individual students to the far more complex and ethically challenging example of "at risk" student predictive modelling and interventions. Future research must include consideration of the student voice to inform learning analytics ethics and privacy debates, as these voices have largely been silent.

REFERENCES

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Educational Technology & Society, 15*(3), 3–26.

Cohen, J. (2000). Examined lives: Informational privacy and the subject as object. *Stanford Law Review, 52*, 1373–1437.

Culver, C., Moor, J., Duerfeldt, W., Kapp, M., & Sullivan, M. (1994). Privacy. *Professional Ethics 3 & 4*, 3–25.

Diaz, V., Golas, J., & Gautsch, S. (2010). Privacy considerations in cloud-based teaching and learning environments. *Educause* (November), 2–10.

Drachsler, H., & Greller, W. (2012). *Confidence in learning analytics*. Paper presented at the LAK12: Second International Conference on Learning Analytics & Knowledge, Vancouver, Canada.

Ellis, C. (2013). Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics. *British Journal of Educational Technology, 44*(4), 662–664.

Floridi, L. (2005). The ontological interpretation of informational privacy. *Ethics and Information Technology, 7*(4), 185–200.

Floridi, L. (2006). Four challenges for a theory of informational privacy. *Ethics and Information Technology, 8*, 109–119.

Fried, C. (1968). Privacy: A moral analysis. *Yale Law Journal, 77*(1), 475–493.

Gerstein, R. (1984). Intimacy and privacy. In F. Shoeman (Ed.), *Philosophical dimensions of privacy: An anthology* (pp. 265–271). Cambridge: Cambridge University Press.

Kang, J. (1998). Information privacy in cyberspace transactions. *Stanford Law Review, 50*(4), 1193–1294.

Moor, J. (2000). Towards a theory of privacy for the information age. In R. M. Baird, R. Ramsower, & S. Rosenbaum (Eds.), *Cyberethics: Moral, social, and legal issues in the computer age* (pp. 200–212). Amherst, NY: Prometheus Books.

Moor, J. (2005). Why we need better ethics for emerging technologies. *Ethics and Information Technology, 7*, 111–119.

Moor, J.H. (1997). Towards a theory of privacy in the information age. *SIGCAS Computers and Society, 27*(3), 27–32. doi: http://doi.acm.org/10.1145/270858.270866

Nissenbaum, H. (2010). *Privacy in Context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford University Press.

Petersen, R., & Worona, S. (2006). Security & privacy: An overview. *Educause* (September/October), 16–17.

Rachels, J. (1975). Why privacy is important. *Philosophy and Public Affairs, 4*(4), 323–333.

Rotenberg, M., & Barnes, K. (2013). Amassing student data and dissipating privacy rights. *Educause* (January/February), 56–57.

Shoemaker, D. (2009). Self-exposure and exposure of the self: Informational privacy and the presentation of identity. *Ethics and Information Technology, 12*(1), 3–15.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1509–1528.

Tavani, H. (1999). Privacy online. *SIGCAS Computers and Society, 29*(4), 11–19. doi: http://doi.acm.org/10.1145/572199.572203

Tavani, H. (2007). *Ethics & technology: Ethical issues in an age of information and communication technology* (2nd ed.). Hoboken, NJ: John Wiley & Sons Inc.

Tavani, H., & Moor, J. (2001). Privacy protection, control of information, and privacy-enhancing technologies. *SIGCAS Computers and Society, 31*(1), 6–11. doi: http://doi.acm.org/10.1145/572277.572278

Tavani, H.T. (2007). Philosophical theories of privacy: Implications for an adequate online privacy policy. *Metaphilosophy, 38*(1).

Warren, S., & Brandeis, L. (1890). The right to privacy (the implicit made explicit). In F.D. Schoeman (Ed.), *Philosophical dimensions of privacy: An anthology* (pp. 193–220). Cambridge, MA: Harvard Law Review.

Westin, A. (1967). *Privacy and freedom*. New York: Atheneum.

Willis, J.E., Campbell, J.P., & Pistilli, M.D. (2013). Ethics, big data, and analytics: A model for application. *Educause Review Online*. Retrieved from http://www.educause.edu/ero/article/ethics-big-data-and-analytics-model-application