Argument complexity: Teaching undergraduates to make better arguments

Matthew A. Kelly & Robert L. West

The task of turning undergrads into academics requires teaching them to reason about the world in a more complex way. We present the Argument Complexity Scale, a tool for analysing the complexity of argumentation, based on the Integrative Complexity and Conceptual Complexity Scales from, respectively, political psychology and personality theory. Argument Complexity classifies arguments based on acknowledgement and consideration of conflicting evidence or conflicting frameworks for judging the issue, use of frameworks for evaluating evidence, and use of meta-frameworks for evaluating frameworks. We discuss how the Argument Complexity Scale can be used to teach undergraduate students to reason and write like academics by providing the scaffolding for forming complex argumentation.

Keywords: *argument complexity; dialectic levels; debate; rationalisation; reasoning skills; writing skills; decision making; intellectual development.*

TO DO RESEARCH, scientists and academics need to understand multiple perspectives on issues, to weigh and assess the evidence, and to communicate this process of reasoning to others in their field. These skills are not just the foundation for science and all of academic argumentation, but also the foundation for all good decision-making, and are critical for ethical and socially responsible decision-making in a diverse and democratic society. Thus, teaching these skills is a critical part of a university education.

Due to the hypothesis confirmation bias, when making an argument, people tend to start with a judgement and then seek evidence in favour of that judgement (Lehman et al., 1992). This is particularly true of moral reasoning, where people typically make judgements on the basis of an emotional reaction, and the reasoning process, if there is one, is then the act of developing a post-hoc justification for their judgement (Haidt, 2001).

A judgement accompanied by supporting evidence is what we will refer to as level one complexity argument. People who already agree with the judgement find level one arguments persuasive. Level one arguments are encouraged in high school English class, where the teachers often instruct students to take a position and present evidence for that position (Nunnally, 1991).

But the Popperian model for science is to try to find evidence *against* judgements (i.e. theories and hypotheses) rather than in favour of them (Popper, 1959). Purely seeking to confirm a position leads to bad science and an unbalanced understanding of the issues. To aid with teaching students the skills they need to be scientists, critical thinkers, and good decision makers, we have developed the Argument Complexity Scale, a theory of argumentation that can be used to both assess the complexity of arguments and to teach students how to reason and argue at an academic level.

In what follows, we review scales of reasoning complexity in the literature and discuss how they differ from our proposed scale. In particular, we detail the Integrative Complexity Scale (Suedfeld & Tetlock, 1977), on which the Argument Complexity Scale is based. We explain the Argument Complexity Scale and provide three examples of arguments analysed using the scale. Finally, we talk about our experiences with using the scale as a teaching tool in a first year undergraduate psychology course.

Our aim in this paper is to introduce the Argument Complexity Scale and describe how the scale can be used as a teaching tool for argumentation in the classroom. While we have had success using the scale in the classroom, the work done so far is preliminary and descriptive. Experimental studies of the validity of the scale in and out of the classroom are a matter for future work.

Complexity scales

Characterising reasoning in terms of different levels of complexity is an old idea. The medieval Islamic scholar Ibn Rushd (known as Averroes in Europe) wrote that there are three ways to religious truth, one for the masses, one for theologians, and one for philosophers (Hillier, 2010). He believed that Islamic truth had to be customised to reach different audiences: The masses should be taught a literal interpretation of the texts, theologians should be taught an allegorical interpretation of the texts, and the philosophers' role was to debate the different possible and conflicting allegorical interpretations.

In Ibn Rushd's analysis we see a historical analogue to the Argument Complexity Scale. Islamic truth as taught to the masses corresponds to the first level of our scale. Believers are taught a perspective on the texts and evidence for that perspective. Islamic truth as interpreted by theologians corresponds to a level 3 on our scale. The theologian is aware of conflicting evidence in the texts and adjudicates the evidence by applying an interpretation to the text. Islamic philosophers debate and adjudicate between different interpretations of Islamic truth, corresponding to levels 4 through 6 on our scale.

Modern scales for measuring complexity include Conceptual Complexity (Schroder et al., 1967), Integrative Complexity (Suedfeld & Tetlock, 1977), Perry's (1970) scheme, and Kohlberg's (1984) stages of moral development. Each of these scales has commonalities with the Argument Complexity Scale, but each scale serves a different purpose and is thus designed differently.

The Argument Complexity Scale is derived from the Integrative Complexity Scale (Suedfeld & Tetlock, 1977), which in turn was adapted from the Conceptual Complexity Scale (Schroder et al., 1967). The Conceptual Complexity Scale measures complexity as personality trait, whereas Argument Complexity measures the complexity of a piece of argumentation. Argument Complexity is not intended to be interpreted as a personality characteristic. While an individual may habitually make arguments at a particular level of complexity, it is also possible for an individual to vary in the complexity of their reasoning depending on the topic.

Integrative Complexity assumes that higher complexity levels entail reasonableness, negotiation, and compromise. Conversely, Argument Complexity does not assume that higher complexity necessitates compromise or adopting a moderate position, but instead characterises higher complexity in terms of the increasingly sophisticated use of evaluative frameworks.

Perry's (1970) scheme describes the typical intellectual and ethical development of students through university. Students transition from black and white thinking to relativism until finally making considered value commitments. Conversely, Argument Complexity characterises an individual's thinking on a specific issue. However, the evaluative frameworks of the Argument Complexity Scale provide a theory for how value commitments arise at the later stages of Perry's scheme.

Kohlberg's (1984) stages of moral development characterise the development of moral reasoning from childhood to adulthood. According to Kohlberg, moral reasoning develops from pre-conventional, to conventional, to post-conventional. Children at the pre-conventional stage make moral decisions with reference to consequences, such as angering their parents. Adolescents and adults at the conventional stage make decisions according to the moral conventions of their social group. Adults at

the post-conventional stage are aware that adhering to convention is not always for the best and may develop moral principles that take precedence over conventions.

When applying Argument Complexity to moral arguments, the scale can resemble Kohlberg's stages. A society's moral conventions are an evaluative framework. Thus, reasoning at the pre-conventional stage (i.e. without a framework) would be at a level 1 complexity, reasoning at the conventional stage could be at a level 3 if the moral conventions are made explicit, and reasoning at the post-conventional stage might be at a level 5 complexity if the moral principles serve as a meta-framework for evaluating contradictory conventions.

The Argument Complexity Scale differs from Kohlberg's stages in that it characterises the development of arguments rather than persons and the scale is not restricted to moral reasoning. Also, as we have found by evaluating the complexity of arguments put forward by first year university students, moral reasoning at the conventional or even the post-conventional stage can be expressed at level 1 on the Argument Complexity Scale. This is because complexity requires reasoners to be aware of and to communicate their evaluative frameworks. It is not enough to present the evaluation without the framework. Moral reasoning, even nuanced moral reasoning, if it occurs without self-awareness of the process of reasoning, is at level 1 complexity. Only through self-awareness of one's own reasoning can reasoning be effectively communicated and properly scrutinised.

Integrative complexity

The Argument Complexity Scale is a modification of the Integrative Complexity Scale developed by Suedfeld and Tetlock (1977). Integrative Complexity is a seven-point scale that measures the complexity of an individual's views on an issue, as evidenced in communications from the individual. Communications with high Integrative Complexity acknowledge different points of view on the issue (*differentiation*) and attempt to combine those differing points of view into a coherent whole (*integration*). Communications with minimal Integrative Complexity expound the individual's opinions without acknowledging the existence of other points of view.

Using Integrative Complexity, the complexity of a communication from an individual is scored according to the amount of differentiation and integration (see Table 1; Baker-Brown et al., 1992). There are four main categories. At level 1, the individual provides an opinion with no differentiation or integration. At level 3, the individual *differentiates* between different perspectives on an issue, but does not integrate those

Table 1: Summary of the integrative complexity scale based on the scoring manual (Baker-Brown et al., 1992).

Level	Critical indicator
1.	*No differentiation*: No mention of conflicting points of view.
2.	*Some differentiation*: Mentions conflicting points of view, but dismisses them.
3.	*Differentiation*: Engages with conflicting points of view.
4.	*Some integration*: Acknowledges conflict between views.
5.	*Integration*: Relates conflicting points of view to each other to achieve a new perspective.
6.	*Some hierarchical integration*: Hints at an overarching view that encompasses all relevant perspectives.
7.	*Hierarchical integration*: Presents an overarching view that integrates perspectives formed through integration.

different perspectives. At level 5, the individual integrates the different perspectives into a single coherent perspective. At level 7, the individual performs hierarchical integration, integrating perspectives that were themselves formed through integration. Even numbered levels 2, 4, and 6 are transitional levels that resemble the immediately lower level with some characteristics of the higher level.

Integrative complexity has been used mainly to analyse political documents. Research has found that the outbreak of war, such as the 2001 invasion of Afghanistan (Suedfeld & Leighton, 2002) or the Gulf War (Wallace et al., 1993), is reliably preceded by a decrease in Integrative Complexity in the communications of national leaders and diplomats. Conversely, crises resolved peacefully typically exhibit rising complexity after a decrease at the initial onset of the crisis (e.g. Maoz & Astorino, 1992; Raphael, 1982; see Suedfeld & Leighton, 2002 for review).

Argument complexity

The Integrative Complexity Scale was designed for analysing negotiations between opposing factions. In such negotiations, compromise is associated with peaceful outcomes. The Integrative Complexity Scale associates higher levels of complexity with both understanding of opposing positions, as indicated by *differentiation*, and compromises reached through *integration* of opposing positions.

Unlike political negotiations, compromise is not necessary for science or good argumentation. The aim of science is (ideally) to find the truth, or, at least, to build evidence for or against the theories of a particular research program (Lakatos, 1976). Science is not concerned with settling disputes amicably between opposing factions.

To suppose that more complex arguments are necessarily more moderate is a form of the *middle ground* or *golden mean* fallacy (Gardner, 2009). The truth, or end of careful reasoning, is not necessarily the middle ground between two opposing positions. The Argument Complexity Scale does not assume that higher levels of complexity approach a middle ground. The scale allows for strong, uncompromising positions to be taken at high (odd-numbered) levels of complexity.

The Argument Complexity Scale incorporates the four main categories of argument identified in the Integrative Complexity Scale, but also includes additional categories. However, the concept of integration is not used in the Argument Complexity Scale. The constructs we use to define the scale are: *differentiation*, *frameworks*, and *meta-frameworks*.

By *differentiation* we mean that a person considers conflicting evidence or conflicting ways of judging the situation. It is possible to differentiate between different reasons for supporting the same cause but this is not what is meant. Differentiation, in our scheme, refers to accepting the legitimacy of conflicting points of view and therefore tolerating some level of ambiguity.

By *frameworks* we mean a frame of reference for judging between alternatives. Judgements always involve a frame of reference; otherwise there is no basis for the judgement. For example, if you think that a particular cake is good, the claim 'This is a good cake' has different meanings depending on the evaluative framework. For example, the judgement of a cake will be different in a cake-tasting event than a cake-decorating event. In daily experience, people are often unaware of the frameworks they are using to judge things. When we refer to the use of frameworks in argument, we mean the explicit, consistent, and deliberate use of frameworks to justify, persuade, and communicate points of view.

We also use the concept of a *meta-framework*. A meta-framework is a high-level framework that is used to judge the value or legitimacy of other frameworks. Often a meta-framework will be more abstract in nature (e.g. religious, moral, analytical or philosophical arguments). For example, utilitarianism may serve as a meta-framework for moral decision-making, the welfare of

a nation's citizens as a meta-framework for political decision-making, and an understanding of the purpose of science as a meta-framework for research decision-making.

What defines a meta-framework is that it is used in the argument to evaluate other *frameworks* rather than to directly evaluate evidence. Whether or not a framework is a meta-framework is not an intrinsic property of the framework, but is instead a property of how the framework is used in a given argument.

For example, suppose you are trying to decide which of two cars to buy. You decide that you care about *cost* and *reliability*. Evaluating the evidence, you find that one car is cheaper and the other car is more reliable. As your two frameworks are in conflict, you do not have an argument for one car over the other. To help you make the decision you can introduce a *meta-framework*.

What framework can serve as a meta-framework for this decision? Let us try using the framework *environmentally friendly*. Suppose that you find that the more reliable car is also more environmentally friendly. While this information might help us make our decision, *environmentally friendly* is not acting as a meta-framework. A meta-framework would evaluate the frameworks of *cost* and *reliability* rather than the directly evaluating the characteristics of the cars. So, *environmentally friendly* is not a meta-framework for the purposes of this argument, rather it is merely an additional framework for evaluating the evidence, like cost and reliability. Conversely, *expected utility*, which accounts for the trade-offs between cost and reliability, could serve as an appropriate meta-framework for the decision of which car to buy.

Levels of argument complexity

In the Argument Complexity Scale, odd numbered levels take a position on an issue, whereas even numbered levels are neutral on the issue. Arguments that are caught somewhere between levels can be indicated using

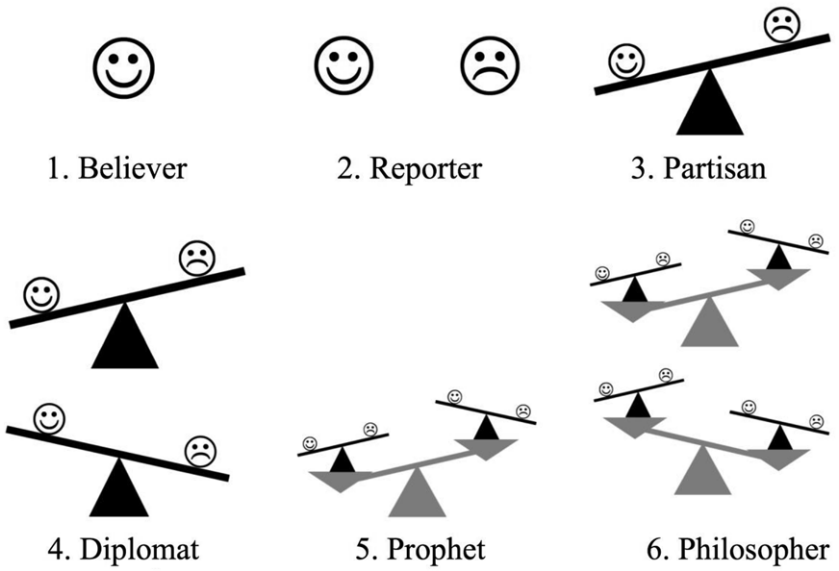

Figure 1: The six levels of the Argument Complexity Scale illustrated. Conflicting points of view are depicted as faces. Conflicting frameworks are depicted as scales that tip in favour of one point of view or the other. Conflicting meta-frameworks are depicted as grey scales that tip in favour of one framework or the other.

fractional levels, e.g. level 1.5. The Argument Complexity Scale has the following levels, which are illustrated visually in Figure 1 and by example in Table 2.

In Table 2, we illustrate the levels of Argument Complexity on three different social issues. At each level, evidence and/or frameworks are used to reach a conclusion. At odd numbered levels, a judgement is made either for or against a proposal (namely, in this set of examples, that abortion should be legal, same-sex marriage should be legal, and that God created humanity). At even numbered levels, judgement is withheld due to conflicting evidence or conflicting frameworks. Table 2 represents judgements

Table 2: Example arguments. Judging for and against are in white and black. Withholding judgement is in grey.

Level	Legality of gay marriage	Legality of abortion	Was humanity created by God?
1. Evidence	Gay marriage is good because it makes gay couples happy.	Abortion should be illegal because it takes the life of an innocent foetus.	According to the Christian Bible, humanity was created by God (Genesis 1:27)
2. Conflicting evidence	Gay marriage is wrong because my pastor told me so.	Banning abortion causes deaths from births or unsafe, illegal abortions.	According to biology, humans and all life evolved from a common ancestor.
3. Framework	***Group rights*** (Jones, 2016), such as the rights of religious groups, are violated by the state imposing a definition of marriage upon them.	Banning abortion conflicts with the right to security of person in the ***Canadian Charter*** (R. v. Morgentaler, 1988).	***Empirical evidence*** from comparative anatomy, fossil records, domesticated plants and animals, etc. support evolution (Darwin, 1872).
4. Conflicting framework	***Individual rights*** are violated when the state interferes with the ability of consenting individuals to marry each other.	***Catholic doctrine*** condemns abortion because life is sacred from the moment of conception (John Paul II, 1995).	Intelligent design appeals to the seeming ***irreducible complexity*** of life to argue that it is must have been designed rather than evolved (Behe, 1996).
5. Meta-framework	***Ethnic nationalism*** holds that nations and the people within them function better when they conform to a shared set of cultural norms, such as a shared language and religion. Legalising gay marriage breaks with tradition, weakening national cohesion.	The purpose of both Canadian law and Catholic doctrine is to ***minimise harm*** to people. Canadian law is better at minimising harm given that it is dubious that a foetus feels pain or thinks.	A theory is only scientific if it is ***falsifiable*** (Popper, 1959), i.e. makes testable predictions, which evolution does and intelligent design does not.
6. Conflicting meta-framework	***Civic nationalism*** holds that nations are defined strictly by citizenship and that nations work best with a code of law that is maximally permissive to allow for cultural and individual diversity.	Under a ***deontological*** view of morality, the purpose of moral rules is to define right action. Taking an innocent human life, even a foetus, is an evil act under this view.	***Non-overlapping magisteria*** (Gould, 1997) holds that religion and science address non-overlapping sets of questions. Thus, *Did humanity evolve?* is a question for science, whereas *Was humanity created by god?* is a question for religion.

for in white, *against* in black, and withholding judgement in *grey*. Arguments at higher levels of complexity are cumulative: they include the evidence and frameworks of lower levels, which are not repeated for the sake of brevity. In Table 2, new frameworks introduced at a particular level of complexity are in ***bold italics***.

Level 1: The believer

The lowest level of Argument Complexity is characterised by an absence of differentiation and no explicit use of frameworks. A level 1 argument presents a position and evidence for that position without discussing alternate positions, contrary evidence, and without explaining why the evidence is relevant. If opposing positions are mentioned, the believer presents them as a straw man argument or dismisses without substantive discussion.

Level 2: The reporter

The *reporter* is characterised by *differentiation*. The reporter is aware of differing points of view on the issue and accepts evidence in favour of those differing points of view. But the reporter does not use judgmental frameworks to choose one side over the other.

Level 3: The partisan

Like the reporter, the *partisan* accepts evidence supporting conflicting points of view of an argument and faithfully presents the conflicting points of view and evidence for those points of view. But the partisan also employs one or more frameworks to justify choosing one side over another. This effectively resolves the ambiguity caused by accepting information from both sides. In the examples in Table 2, conflicting evidence on abortion is resolved by appeal to the framework of the *Canadian Charter of Rights and Freedoms*. Likewise, the debate between creationism and evolution is resolved by appeal to *empirical evidence*, and the same-sex marriage example introduces the framework *group rights* (Jones, 2016).

Level 4: The diplomat

The *diplomat* is characterised by differentiation in terms of accepting evidence in support of conflicting points of view and also differentiation at the framework level in terms of understanding that different, conflicting frameworks can be used to judge the evidence to be in favour of different sides. In the abortion example, the framework of *Catholic doctrine* (John Paul II, 1995) conflicts with the *Canadian Charter of Rights and Freedoms* (R. v. Morgentaler, 1988). The diplomat accepts the validity of both frameworks and tolerates both points of view. Compromise and novel solutions may be offered at this level, which would correspond to *integration* in the Integrative Complexity Scale (Suedfeld & Tetlock, 1977). However, a level 4 argument does not necessarily offer a compromise or solution to the conflict.

Level 5: The prophet

Like the *diplomat*, the *prophet* accepts evidence in support of conflicting points of view and understands that different, conflicting frameworks can be used to judge the evidence to be in favour of differing point of view. The prophet is characterised by the use of a meta-framework (or meta-frameworks) to adjudicate which frameworks are the right ones to use when evaluating the evidence. The meta-framework is *not* used to make a judgement on the issue directly, but is instead used to judge which is the correct framework. At this level, ambiguity over conflicting judgemental frameworks is resolved through the use of the meta-framework, allowing the prophet to reach a conclusion in favour of a particular point of view.

In the abortion example in Table 2, the principle of *minimising harm* from the utilitarian theory of ethics is used as a meta-framework to judge which of the frameworks, the *Canadian Charter of Rights and Freedoms* or *Catholic doctrine*, is the better framework to use for deciding the abortion issue. In the same-sex marriage example, *ethnic nationalism* is used to judge that group rights should take priority over individual rights.

In the creationism example, the framework used to justify intelligent design (*irreducible complexity*) implies that intelligent design is intended as a scientific theory. As such, it is appropriate to resolve the conflict between the theories and frameworks in this example by appeal to philosophy of science. Thus we invoke *falsifiability* (Popper, 1959) to adjudicate in favour of evolution.

Level 6: The philosopher

The *philosopher* is characterised by accepting evidence in support of conflicting points of view, differentiating conflicting frameworks for judging the evidence, and also *differentiation at the meta-framework level*. The philosopher understands that different, conflicting meta-frameworks can be used to judge which frameworks are relevant to the task of evaluating the evidence, and understands that the conclusion reached is contingent on the choice of meta-framework.

In Table 2, for the gay marriage example, *civic nationalism* is a conflicting meta-framework for *ethnic nationalism*. In the abortion example, *deontological* ethics conflicts with *utilitarian* ethics. In the creationism example, the role of the conflicting meta-framework *non-overlapping magisteria* (Gould, 1997) is less straightforward. *Non-overlapping magisteria* adjudicates *against* all lower level frameworks. If science and religion are *non-overlapping magisteria*, then the *empirical evidence* and *irreducible complexity* frameworks are both irrelevant as 'Was humanity created by God?' is a religious rather than a scientific question and so must be addressed by religion rather than science.

Level 7 and upward

There is not, in principle, an upper limit to the Argument Complexity Scale. A level 7 on the scale would apply a meta-meta-framework to adjudicate between meta-frameworks. A level 8 would differentiate between conflicting meta-meta-frameworks. A level 9 would apply a meta-meta-meta-framework, and so on. However, arguments with higher levels of complexity are more difficult for people to produce or understand and tend to be long, costing greater amounts of time, energy, and cognitive processing. As a result, arguments of lower complexity are more common. Levels 1 to 6 cover most arguments encountered in our everyday lives, from the dinner table, to newspapers, to academia.

There may also be logical limits to our ability to generate higher-level frameworks to adjudicate between frameworks. As a child that repeatedly asks 'why?' discovers, there is a point when we run out of frameworks for justification. For example, for the abortion argument in Table 2, a Level 7 argument would require a meta-meta-framework for judging one theory of normative ethics to be better than another (utilitarianism versus deontological ethics). But judging one theory of normative ethics to be more moral than another would seem to require appeal to a theory of normative ethics, which is circular. Thus the choice of a theory of normative ethics seems, by necessity, an arbitrary one. That is to say, there may not be an appropriate meta-meta-framework.

Conversely, in the creationism example in Table 2, we can continue to level 7. The conflicting meta-framework, *non-overlapping magisteria,* is also an appropriate *meta-meta-framework* as it can be used to evaluate the conflicting meta-framework, *falsifiability*, as appropriate for scientific questions, but not for religious questions. It is worth noting that more complex arguments are not necessarily more correct. For example, *non-overlapping magisteria* is a controversial position (Dawkins, 2006).

Teaching

Argument Complexity was developed to teach better essay writing skills to first year undergraduates. For the past several years we have been asking first year undergraduates taking a first year cognitive science course to write an essay on an issue they feel strongly about. After handing it in they receive a lecture on the complexity scale and are instructed to re-write the essay at a higher level of complexity. To motivate them, in

previous years, the re-written essay has been marked based on their complexity score. Disturbingly, the essays before the lecture were almost all at level 1 and the majority of the follow up essays failed to reach level 3, and no student ever achieved a level above 3 on the Argument Complexity Scale.

We can speculate as to why undergraduates might have difficulty learning to argue at a higher level of complexity. Complexity could be a personality trait, such that individuals have a preferred level of complexity at which they habitually reason (Schroder et al., 1967).

Alternatively, the way essay writing is taught in high school may encourage writing level 1 arguments. A standard method of teaching essay writing, known as the five-paragraph theme (Nunnally, 1991), is to get students to pick a position and then present three pieces of evidence in favour of that position. This method produces level 1 essays as *differentiation* and identification of *frameworks* are not requirements.

Another possible reason for why undergraduates might have difficulty with higher levels of complexity is that popular media and politicians overwhelmingly present arguments at a level 1 complexity, which can create a culture in which level 1 arguments are the norm. In a review of political speeches and media coverage of Canadian elections over the decades, Suedfeld et al. (1990) found that all media and politicians had mean integrative complexity scores of less than 3, which corresponds to mean argument complexity scores of less than 2.

To see if more extensive training on complexity could make a difference, we extended teaching complexity over the term. We had two hour and a half workshop classes that provided students with the opportunity to work in groups to practice explicitly identifying the frameworks and meta-frameworks that they use to make judgements. The class had 23 students that participated (10 male, 13 female). Students completed four assignments. For each assignment, they were asked to write an argument. The complexity of their arguments is shown in Figure 2.

For the first assignment, students were asked to present an argument on a controversial issue of their choice. Almost all of the students wrote a level 1 complexity argument. After the first assignment, students were taught the Argument Complexity Scale. For the second assignment, students were

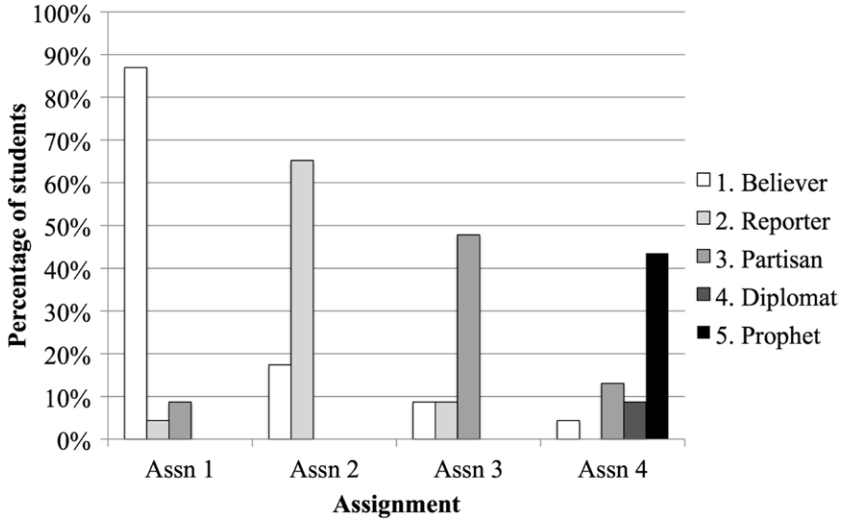

Figure 2: Complexity levels of student assignments.

asked to present the opposing position to the argument they presented in the first assignment in order to achieve a level 2 complexity. Almost all of the students were able to do this, but a few were unable to conceptualise the opposing point of view. Presenting evidence for the opposite point of view can be an aversive experience for students who feel strongly about the issue upon which they have chosen to write.

For the third assignment, students had to identify a framework for adjudicating between the opposing positions they described in the first and second assignments. Identifying frameworks was still difficult for the students, as it requires not just reasoning ability, but a self-awareness about how you are reasoning, and the ability to characterise your own reasoning process. Identifying frameworks requires introspection and a degree of abstraction that first years are likely to have had little experience with in their education up to this point.

For the fourth assignment, students were assigned a different topic than the first three assignments and had to identify opposing positions and evidence, opposing frameworks, and a meta-framework for judging between the opposing frameworks. We had students write the fourth assignment argument in five steps to help them reach an Argument Complexity level of 5:

1. Present evidence for one side of an issue.
2. Present evidence for the opposite position.
3. Identify a framework that supports one side.
4. Identify a framework that supports the other side.
5. Identify a meta-framework for judging between the two conflicting frameworks.

This highly structured assignment was successful and most students who completed the assignment produced a level 5 document. Our experiences with using the Argument Complexity Scale as a teaching tool show that:

1. First-year undergrads habitually write arguments at a level 1 complexity.
2. With practice and appropriate scaffolding, first-year undergrads can write arguments at a level 5 complexity.

These observations suggest that the Argument Complexity Scale is useful both as analysis tool and as a pedagogical tool. By explicitly teaching the Argument Complexity Scale to undergraduates, we may be able to accelerate the process of learning to write at the level of complexity expected in academia and needed in the world. Sadly, without any explicit training, many students will obtain an undergraduate degree without moving beyond level 1.

Future work

Work still needs to be done to evaluate the validity of the Argument Complexity Scale and its usefulness as a teaching tool. Does the Argument Complexity Scale actually align with academics' implicit expectations for good arguments? The validity of the scale could be evaluated by having professors or teaching assistants mark essays written at varying levels of complexity. By holding the quality of writing and reasoning constant across the essays, one could empirically demonstrate if there is a positive correlation between grades assigned and complexity of argumentation as measured by the scale.

Do the expectations for complexity in argumentation vary from one discipline to another, or do academics favour more complex argumentation consistently across disciplines? Philosophy, for example, might expect a higher degree of complexity in argumentation than some (or all) of the sciences. The proposed essay marking study may be worth conducting across multiple departments to investigate the possibility that some disciplines have a stronger preference for more complex arguments.

Future classroom studies will be important for establishing the pedagogical use of the Argument Complexity Scale. However,

the scale touches on only one aspect of good argumentation: the complexity of an argument. Clear communication and careful reasoning are also important to good arguments, and students enter the classroom with varying abilities to communicate and think critically. Because of the noise introduced by this spread of skills, it may prove difficult to test the effect of teaching the Argument Complexity Scale on the ability of students to make good arguments, particularly with small class sizes.

Conclusion

We introduce the Argument Complexity Scale, a theory and measure of the complexity of argumentation. The Argument Complexity Scale can be used to teach undergraduate students to reason at the level required by academia by providing an explicit analysis of what academic argumentation consists of and the scaffolding they need to produce complex argumentation.

All else being equal, a more complex argument is a better argument than a less complex argument. Complexity adds more to the understanding of an issue without taking anything away. However, the Argument Complexity Scale provides only a high level framework for good argumentation. Clarity of communication and carefulness of reasoning are other important aspects of constructing a good argument. A complex but poorly reasoned or poorly communicated argument is still a bad argument. In teaching students good argumentation, communication and reasoning skills need to be fostered alongside an understanding of the importance of complexity.

In this paper, our aim was to outline how the Argument Complexity Scale can be used in the classroom. Our approach is to teach the scale as a theory of argumentation and then follow up that lecture with exercises for identifying frameworks and writing arguments at varying levels of complexity. Teaching students to be self-aware about the evaluative frameworks they use when reasoning is the most difficult part of teaching the Argument Complexity Scale. Meta-frameworks are particularly difficult for students to identify. In light of this, it may worthwhile to teach the scale alongside lectures on how people make decisions. For example, Moral Foundations Theory (Haidt & Joseph, 2007) proposes that five moral foundations (*care*, *fairness*, *loyalty*, *respect*, and *purity*) serve as the frameworks or meta-frameworks underlying all moral decision-making. Providing examples of common frameworks is a critical part of teaching students how to identify their own frameworks and effectively use the Argument Complexity Scale.

The authors

Matthew A. Kelly, Post-doctoral Researcher, Applied Cognitive Science Lab, College of Information Sciences and Technology, Pennsylvania State University, E397 Westgate Building, University Park, PA 16802 - 6823.
Robert L. West, Associate Professor, Carleton Cognitive Modeling Lab, Institute of Cognitive Science, Carleton University, 2202A Dunton Tower, 1125 Colonel By Drive, Ottawa, ON K1S 5B6.

Correspondence

Matthew A. Kelly, College of Information Sciences and Technology, Pennsylvania State University, University Park, PA.
Email: matthew.kelly@psu.edu

References

Baker-Brown, G., Ballard, E.J., Bluck, S., de Vries, B., Suedfeld, P. & Tetlock, P.E. (1992). The conceptual/integrative complexity scoring manual. In C.P. Smith, J.W. Atkinson, D.C. McClelland & J. Veroff (Eds.) *Motivation and Personality: Handbook of thematic content analysis* (pp.401–418). New York: Cambridge University Press.

Behe, M.J. (1996). *Darwin's black box: The biochemical challenge to evolution*. New York: Simon and Schuster.

Darwin, C.R. (1872). *The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life* (6th edn). London: John Murray.

Dawkins, R. (2006). *The God Delusion*. London: Black Swan. ISBN 978-0-552-77331-7.

Gardner, S.T. (2009). *Thinking your way to freedom: A guide to owning your own practical reasoning*. Philadelphia: Temple University Press.

Gould, S.J. (1997). Nonoverlapping magisteria. *Natural History, 106*, 16–22.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

Haidt, J. & Joseph, C. (2007). The moral mind: How 5 sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence & S. Stich (Eds.) *The innate mind* (Vol. 3, pp.367–391). New York: Oxford University Press.

Hillier, H.C. (2010). Ibn Rushed (Averroes) (1126–1198). Internet Encyclopedia of Philosophy, ISSN 2161-0002. Retrieved 3 February 2015. http://www.iep.utm.edu/ibnrushd/

John Paul II (1995). *Evangelium Vitae*, 61. Retrieved 12 April 2016. http://w2.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_25031995_evangelium-vitae.html

Jones, P. (2016). Group rights. In E.N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy* (Summer 2016 edn). Retrieved 14 April 2016. http://plato.stanford.edu/archives/sum2016/entries/rights-group/

Kohlberg, L. (1984). *The psychology of moral development: The nature and validity of moral stages* (Essays on Moral Development, Vol. 2). San Francisco: Harper & Row.

Lakatos, I. (1976). Falsification and the methodology of scientific research programmes. In S.G. Harding (Ed.) *Can theories be refuted?* (pp.205–259). Netherlands: Springer.

Lehman, D.R., Krosnick, J.A., West, R.L. & Li, F. (1992). The focus of judgment effect: A question wording effect due to hypothesis confirmation bias. *Personality and Social Psychology Bulletin, 18*, 690–699.

Maoz, Z. & Astorino, A. (1992). The cognitive structure of peacemaking: Egypt and Israel, 1970–1978. *Political Psychology, 13*, 647–662.

Nunnally, T.E. (1991). Breaking the Five-Paragraph-Theme Barrier. *The English Journal, 80*, 67–71. doi:10.2307/818100

Perry, W.G., Jr. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart, and Winston.

Popper, K.R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Raphael, T.D. (1982). Integrative complexity theory and forecasting international crises: Berlin 1946–1962. *Journal of Conflict Resolution, 26*, 423–450.

R. v. Morgentaler, [1988] 1 S.C.R. 30. (1988).

Schroder, H.M., Driver, M.J. & Streufert, S. (1967). *Human information processing*. New York: Holt, Rinehart, & Winston.

Suedfeld, P., Bluck, S., Ballard, E.J. & Baker-Brown, G. (1990). Canadian federal elections: Motive profiles and integrative complexity in political speeches and popular media. *Canadian Journal of Behavioural Science, 22*, 26–36. http://10.1037/h0078936

Suedfeld, P. & Leighton, D.C. (2002). Early communications in the war against terrorism: An integrative complexity analysis. *Political Psychology, 23*, 585–599.

Suedfeld, P. & Tetlock, P.E. (1977). Integrative complexity of communications in international crises. *Journal of Conflict Resolution, 21*, 169–184.

Wallace, M.D., Suedfeld, P. & Thachuk, K.L. (1993). Political rhetoric of leaders under stress: Findings from the Gulf crisis. *Journal of Conflict Resolution, 37*, 94–107.