On the Possibility of Interventions Aimed at Improving Character

Tone Kvernbekk

These are unpublished conference papers for the 'Can Virtue Be Measured?, held by the Jubilee Centre for Character and Values at Oriel College, Thursday 9th – Saturday 11th January 2014. These papers are works in progress and should not be cited without author's prior permission.

School of Education

University of Birmingham

Edgbaston

Birmingham

On the Possibility of Interventions Aimed at Improving Character

Tone Kvernbekk, University of Oslo
tone.kvernbekk@iped.uio.no

The development of virtues and character have always been considered worthy, overall aims of education. Such overall aims might however recede into the background and become inert or paid lip-service to only. The current educational conceptual landscape is dominated by a vocabulary of learning outcomes, testing, measurement, qualification, employability, accountability, effectiveness, competencies and predictability; here given in no particular order. Since schooling and education are considered successful when predetermined outcomes have been achieved, education makes excessive requirements of assessment, measurement and documentation. Overall aims, as opposed to outcomes or objectives, are long-term aims. They are often vaguely stated and exceedingly hard to operationalize (just consider the term *Bildung*), and thus seemingly do not fit the current landscape very well.

Now there seems to be a rekindled interest in virtue enhancement and character improvement, albeit it within the framework of the landscape outlined above. Improvement of character has become a matter of evidence-based practice (EBP): we devise programs and conduct interventions in order to improve character. Various stake-holders, such as teachers, parents and policy-makers, wish to know whether such programs are effective; that is, whether or to what degree they *work*. It is the business of EBP to identify what evidence we need to make good judgments about the effectiveness of our (proposed) interventions. Indeed, it seems fair to say that one of the ambitions of EBP is that our knowledge, which is evidence-based, should allow us to predict with some reliability that use of a method (program, strategy) will bring about the goal we wish to achieve. In other words, we find out *what works*.

I shall in this paper throw doubt on the idea that character improvement (or virtue enhancement) lends itself to interventions understood within an EBP framework. My discussion will follow two different pathways. First I shall focus on the logic of intervention. EBP is a huge domain and this is but one aspect of it; nevertheless this one aspect requires us to cover quite a lot of ground. I shall argue that it is in general harder and more demanding than we think (or like, if we are advocates of EBP) to make an intervention work, but it is not impossible. Second, I shall argue that one cannot teach virtues the way one teaches maths or languages to students. Virtue enhancement and character formation belong to a different logical type. They are better understood as results of second-order learning or context learning

and do not, therefore, lend themselves to interventions. This understanding of character stems from system theory, which forms quite a contrast to intervention theory. Let me first delineate very briefly the terms *virtue* and *character*. I am going to make do with thin definitions. By a virtue I am simply going to mean a good, admirable or desirable personal quality, and by character I am going to mean the aggregate of virtues, features and traits that form an individual's nature or personality. Obviously any content can be given to these traits, but particular manifestations of virtue and character do not matter to my analysis and so I prefer to keep it open.

The logic of intervention

Roughly speaking, an intervention is a set of actions we perform because we want to change something. More precisely, we implement a causal claim which says that method (policy, strategy, program) X leads to or contributes to effect (result, learning outcome) Y. An intervention we say is effective to the extent it contributes to the desired change, and EBP provides evidence concerning this effectiveness. This may seem simple enough, but once we start digging into it complexities emerge to confound the picture.

EBP is causal in nature, as are interventions. Causality comprises a vast, complex domain with many competing causal theories and concepts. We have regularity theories, counterfactual, probabilistic and intervention theories as well as process theories and capacity approaches. The causal vocabulary contains such concepts as necessity, sufficiency, omissions, absences, prevention, pre-emption, inus-conditions, enablers and disablers in addition to the basic concepts of cause and effect. These can be combined in various ways and are to larger or lesser degrees accommodated by the different theories. There are many possible inroads into this nexus of problems, and I choose to look at certain aspects concerning interventions.

In our case here we wish to intervene in order to enhance virtues or improve character. We must assume that students already have both virtues and character, we are just not entirely happy with the values of those "variables", so we devise a program to enhance them. The already existing value might be termed the *default value* (Hitchcock 2007), the value we would expect a variable to have in the absence of intervening causes. The default, Hitchcock underlines, is not that the end state in question *is* this or that, but that it will *remain* this or that unless we do something or something happens. When a set of variables all take on their default value – business is run as usual, we might say – they cannot by themselves cause

another variable to take a different, deviant, value. Hitchcock sees this as a natural principle of causal reasoning. If a variable should take on a deviant (or unexpected) value, there must be some outside variable or event that explains it. That is, to change the value of our target variable (character formation) we have to intervene somehow.

It is not obvious that there is a clear conception of intervention in EBP, used and understood by advocates, practitioners and critics alike. As far as I can see, this particular topic is not much discussed in the otherwise burgeoning EBP literature. On the other hand, it seems that many different assumptions about causality are made and it is relevant to my topic here to tease out some of them. First, at the very least EBP requires an understanding of causality that accommodates human actions as causes, and that brings us to manipulationist and interventionist theories (Woodward 2008). There must be some factors that lie within the domain of the educators to manipulate. 'Manipulation' is here a technical term and simply means to change (the value of) input variables in order to change the output. On this type of causal theory, to say that "X leads to Y" is to say that if I intervene on X (input variable) and manipulate its value, this will lead to or contribute to a change in Y (output). We intervene on the causes and not on the effect – character improvement in students (or others) is not directly fashioned by teachers, parents, or others. Yet manipulationist/interventionist theories capture various psychological insights we have about our own role in causal learning. We act, we intervene, we create changes, we observe the consequences and we draw inferences about how things hang together and what we should do to achieve the results we desire. In our case here, input X is a program which might consist of different elements (e.g. selected texts, films, discussions, exercises, role-plays) and output Y should be enhanced virtues or an improvement of character. The teacher implements X, Y resides with the students.

To speak in terms of X and Y may seem like an illegitimate over-simplification, since we know that causal connections can be hugely complex and manifold. Nevertheless, this seems to be what we are after: "large" overall relations between X and Y. What we strive to find, in EBP and in practical pedagogy in general, are regular relations between input and output which (we hope) allow us to predict the output with some confidence and thus plan our actions and classroom activities. It should be noted that in cases like this one, both input and output surely are complex entities. Programs can consist of many elements, and character improvement surely has several sources and many facets.

Causal presuppositions

I shall look specifically at three closely inter-related characteristics of the X-Y relation that I believe are simply assumed by both advocates and critics of EBP; that such causal relations are *sufficient*, that they are *stable*, and that they are somehow *basic* in that they operate independently of the structure of the world. I have singled out these three for special attention not only because they are tacitly assumed in the educational EBP literature, but because they seem to be essential to the whole EBP enterprise and because they will figure prominently in my problematization of whether character improvement lends itself to interventions and EBP.

Let us begin with sufficiency. It is important for any researcher who comes up with, say, a new teaching method to be able to say that the method is sufficient for the result. Very often the vocabulary used reveals it: *will lead to, will promote change, will produce, will bring about* – these are all typical expressions of causal sufficiency. A sufficient cause is a cause that brings about its effect unfailingly. It may be difficult bordering on impossible to uphold any claim to sufficiency, but it is also vital to do just that, for several reasons. First, sufficiency is intimately connected to manipulation. It is precisely the sufficient causes that can be manipulated to bring about a change in the output – manipulation of *necessary* causes will not in and of itself bring about changes in output. Second, sufficient causes are surely direct rather than indirect. If a cause is indirect, its effect is modified in more or less predictable ways by mediating factors and we may not rely on them leading to the effect. Third, causal sufficiency is connected to the possibility of prediction and thus to the possibility of planning for achievement of certain results. Knowing that X is sufficient for Y, in order to achieve Y we should do X – on the assumption that the cause is direct and not mediated by other factors which might dilute its effect. Causal knowledge offers possibilities for prediction, foresight and planning, all very useful in professional contexts as well as in daily life. Finally, causal sufficiency seems intimately tied to generalization, since sufficient connections are thought to be invariant and therefore eminently transferable to different contexts.

Causal stability naturally belongs in a cluster already consisting of manipulability, sufficiency, predictability and generality, simply because only stable relationships are manipulable. Relations that are unstable, random, elusive or extremely low-frequent are exploitable neither for planning nor for explanation and predictable achievement of results. The general idea that something *works* is predicated on stability as well as sufficiency. The evidence that we rely upon in EBP, which according to the orthodoxy should come from

randomized controlled trials (RCT), indeed tells you about a stable relation between cause and effect, input and output. Stability of the causal X-Y relation is a natural assumption to make given the generality of many educational programs. They have been shown, by RCT studies, to work in many places. The cause does not produce its effect by accident, but stably and reliably. And that brings me to the third assumed characteristic; that such stable, sufficient, general causal relations are somehow *basic*. We seem to assume that once RCTs have been conducted to show that the relation in question holds in several different places, we can safely conclude that this relation simply *exists* and that knowledge of it can be employed to produce the same effect elsewhere. That is to say, we assume the relation holds *simpliciter*, in and of itself. Such relations are surely eminently generalizable. Sufficiency, stability, generality, predictability and *simpliciter* are assumptions that go well together and probably mutually reinforce each other.

This story might express the hopes of many politicians and EBP advocates. If it were true, we could perhaps devise programs that would allow us to enhance the virtues and improve the character of students everywhere – assuming, of course, that we could agree on what the desirable virtues and character are, and assuming that we could measure them. In passing, I would think that virtues are measurable; it depends on what one takes measurement to be. Assignment of numerals is not essential to measurement; the problem is rather to find good indicants.

The importance of pre-existing conditions

RCT evidence tells us that a causal relation has worked *somewhere*, perhaps that it has worked in many places. But it does not tell us whether the regularity will bring about the desired result in "our" school, should we implement it. Interventions are more complicated than that. Let us go back to Christopher Hitchcock's (2007) distinction between default and deviant values of variables. The importance of this distinction is that it indicates the existence of what we might call a *causal system*. This system already produces an output, the default value, and unless we intervene the value will remain the same. Any intervention is thus inserted into pre-existing conditions, conditions that are often ignored by both program architects and practitioners. The hypothesis is, as stated above, that if we deliver program X, it will bring about result Y or some improved Y. But how are such systems to be understood? Causation theorists who advocate manipulationist/interventionist theories, for example Judea Pearl (2001, 2009) basically sees causal systems as being made up of autonomous causal

mechanisms; represented by equations in which the value of one variable is causally determined by the values of others. On this view, when we intervene we lift X (the cause) from the influence of its old mechanism and place it under the influence of a new mechanism. That is to say, we re-set the value of X from x1 to x2 while keeping all other mechanisms undisturbed. Any other variables that change do so only if they are effects of X. What we do, according to Pearl, is disconnect the intervened-on variable(s) from its normal causal history. Interventions disrupt the relationship between X and its parents, such that the value of X now is determined by the intervention – that is, by us. X is severed off from its normal factors of influence; it overrides them and renders them irrelevant (Sloman 2005). X (the input factors that contribute to character formation) is no longer sensitive to changes in other factors but depends only on our intervention. We have made a shift from one causal path to another. X is disconnected from its usual antecedents so the default assumption that the system will continue to produce its usual output is no longer valid. Rather, with a changed value of X and a new causal path, we predict (expect, hope…) that the output will change in desirable ways. It should be noted that the intervention itself thus stands in a causal relationship to X. This way of understanding interventions is highly useful when we try to predict what will happen to Y (our target) if we intervene. Pearl requires that all other mechanisms are left intact, and we can therefore concentrate on the effects of x1 exclusively and evaluate what would happen if the value changed, and sometimes we do want to do that.

I can see why this way of understanding intervention might be attractive. We can intervene "surgically" on one variable and have an equally specific change at the effect end, without disturbing the other factors in the system. The other mechanisms are left intact and we need therefore not worry about them, which surely is a good thing when we decide whether to implement an intervention or not. There is, however, disagreement about the nature of the causal X-Y connection here – Pearl (2001) does say it is invariant (sufficient) and is criticized for that by Nancy Cartwright (2009), whereas Hitchcock (2001) says that Pearl's view does not entail that interventions are miraculous; mechanisms are not inviolable laws but rather *ceteris paribus* laws that can be disrupted by external interventions. Like I said above, I think the prospect of invariance (sufficiency) is alluring to politicians, school bureaucrats and practitioners alike; since sufficient causes do bring about their effects stably and predictably. I shall come back to the issue of basicness shortly.

There are a couple of problems with this conception of intervention, enticing though it might be. First, can we assume that causal mechanisms are autonomous? It seems to me we cannot – educational systems must rather be understood as open systems, complex entities

where things hang together in various ways, interact, and influence each other in various ways and to different degrees, and with room for random occurrences. Character formation and improvement have many sources. The interventionist theory says that Y will be produced (or the probability of Y will increase) if we do X only if among the possible causes of Y we change solely X (and its immediate "downstream" effects), so that the system stays the same. But does it not stand to reason to think that when we intervene to change X, or we introduce X into the system, we generally change a host of other factors as well, not just the ones that are causally "downstream" from X? If X takes time away from other potentially character-building activities, then our intervention might – albeit unintentionally – prevent other factors from contributing positively to Y. Intervening on X may inadvertently change other causal connections that have Y in their consequent, for better or worse.

But the problems do not end there. Causal relations, Nancy Cartwright (e.g. 2001, 2012) says, are not basic. They *local* and hold only *ceteris paribus*. They do not hold *simpliciter* and are neither stable nor sufficient nor general. To see how she comes to go against the current of interventionist theory, we need to look at the relationship between a system and the causal connections we are interested in. This relationship, Cartwright says, is very close. We have to distinguish here between causal regularities and underlying causal structures. Her basic idea is that causal regularities, such as the relation between an intervention and its effect, are completely dependent upon an underlying causal structure. They cannot even *arise* but for such a causal structure, which I have here loosely termed a 'system'. The structure is primary and the causal regularity is secondary. Regularities are therefore not basic, as so much EBP literature assumes (both advocates and critics alike). They depend on the existing, underlying causal structure, not on external interventions. And here a further complication enters the picture: ontology. Nancy Cartwright is a pluralist. The world is deeply manifold and unruly and causal structures found in different places are rarely if ever causally homogeneous: "Outside the supervision of a laboratory or the closed casement of a factory-made module, what happens in one instance is rarely a guide to what will happen in others" (2001, p.86). Two presuppositions for the generalizability of causal relations are thus instantly problematized: causal stability and causal homogeneity. We cannot assume that the causal structures *here* and *there* are common. While the structure may support X there, it may not do so here. And without some form of similarity, the influence of the cause cannot travel to the effect here, it has no route to take, there is no supporting causal structure:

> Because so many of the causal principles [regularities] we employ are tied to causal structures that underpin them, you can't just take a causal principle that applies here, no matter how sure you are of it, and suppose it will apply there. After all, common causal structures are not that typical, even in the limited and highly controlled world of structures we engineer (2012, p.978).

If Cartwright is correct (and she argues convincingly for her views), then any causal regularity depends on the "right" kind of causal structure to be able to do its work. An intervention that works well in one school may not work in another school, simply because the local structure does not support it.

This does not mean that planning and foresight are impossible. The causal relation between intervention and outcome is local, not general. If an RCT shows that the relation holds *somewhere*, it means that the right kind of causal mix was present for X to find its way to Y. The RCT does not tell us what will happen if we introduce X into a different structure to bring about Y (contrary to what I think most researchers and practitioners believe about RCT evidence, namely that the causal relation holds *simpliciter*). The stability we are after for planning purposes, resides in the underlying structure and not in the causal relation. But even causal structures are a changing mix of factors; consist as they do of individual preferences, relationships of friendship and conflict, norms, written and unwritten rules, values, curricula, parental involvement, administrative and material resources, patterns of communication, etc. It is into such systems that we insert an intervention. The system already produces an outcome, we assume, and that means that there already is at least one route, in all likelihood several, from some (perhaps unknown) input factors to character formation, which is our desired effect Y. Of all the variables present in a complex system, we privilege the one that is our intervention, which we think we can manipulate. If we ask ourselves whether we should try this program for character improvement, we foreground the "this" (Cartwright and Hardie 2012). This form of salience is natural, and it matters greatly because it brings neglect of other factors in the pre-existing structure. Some of the neglected factors and relations may be necessary for the intervention to work at all; that is, for the underlying structure to provide paths so the cause can travel to its effect. The salience might also make us blind to the default running of the system and therefore to the possible ways in which our intervention might hinder already existing causes of Y from making their effect happen. If we are unlucky, implementing the intervention undermines the structure that gives rise to the regularities we rely on to predict the outcomes of our intervention. The pre-existing structure may not stay

fixed under intervention, but of course – its changes may both facilitate and hinder X from contributing to Y.

Summing up

I hope my arguments make clear that while EBP is not impossible, it is more difficult than generally assumed. The causal claims warranted by RCT evidence (prescribed by the EBP orthodoxy) are rather restrained due to the trade-off between internal and external validity. RCTs may establish with a high degree of certainty that input causes output in the population studied and under the conditions under which the input is administered. Results have external validity if they hold for other populations under other circumstances, but this we simply cannot assume. The world is too diverse, local structures (contexts) differ and we cannot presuppose that they are causally homogeneous. What works here might therefore not work there. Furthermore, the privileging of our program variable (which we can manipulate) takes attention away from other factors in the system into which X is inserted, and some of these might be necessary for our intervention to work. Incidentally implementing X may also mean thwarting already existing factors from contributing to Y. On the other hand it may also enhance their contribution, which makes it difficult to evaluate to what extent X actually contributed to Y. And of course, we face problems of measurement, and the problem of justifying why some observed change should count as an *improvement* and not just a change. Daunting tasks, but not impossible.

Character: a case of context learning

In the previous section I have argued that evidence-based interventions generally are much more complicated than we think. The difficulties have different sources; the diversity of the world, the limited relevance of RCT evidence, various misleading causal presuppositions, and of course special problems pertaining to the kind of intervention we are interested in here, namely character improvement. I do conclude, however, that while EBP is more demanding than recognized, it is not impossible. In this section I shall argue that character improvement does not lend itself to intervention; it is simply the wrong kind of entity for the EBP enterprise.

In exploring this line of reasoning I will employ system theory. This forms quite a contrast to the logic of intervention. I do not, however, think the two perspectives are downright contradictory – and I hope not, since I intend to bring insights from system theory

to bear upon the logic of intervention. To begin with, we meet a very different vocabulary. Gregory Bateson, my main theorist, deals in systems, patterns, order, flexibility, hierarchies, logical types, context, difference, self-correction, communication, information, punctuation, change, etc. (Bateson 1972, 1980, 2009). From Ludwig van Bertalanffy, a system theory classic, we get some of the same plus an additional emphasis on organizational feedback, wholeness, directedness, etc. (Bertalanffy 1984). From William Ross Ashby, a pioneer in the closely related field of cybernetics, we get input, output, transformation, control, circuits, restraint, stability, etc. (Ashby 1964). Taken together they make up a way of thinking that deals with form rather than substance and that carves up the world in different and unexpected ways. Like any vocabulary, this one encourages certain ways of understanding phenomena and precludes others. A consistently Batesonian way of thinking about interventions will constantly rub against customary, traditional vocabularies and deeply ingrained ways of thinking about the world.

Some basic systemic tenets

A system, von Bertalanffy says (1984, p.19), is organized complexity, made up of parts in interaction. *Complexity* has to do with the number of parts involved and their internal patterns, and *organization* refers to the level of order/disorder in the system. Some causal theorists, for example Nancy Cartwright and Christopher Hitchcock, also think in terms of (causal) systems when they discuss intervention, whereas Judea Pearl, as we have seen, thinks in terms of autonomous causal relations that can be "surgically" intervened on without disturbing other factors. Conventional systemic wisdom tells us that causal mechanisms (as Pearl calls the autonomous X-Y relations) cannot be changed without affecting other parts in the system. But this is not really the crux about systems, as I understand Bateson. Yes, they do hang together internally. But more importantly, systems are part-whole relations, hierarchically organized. In a much-cited passage he says,

> … I speak of an action or utterance as occurring "in" a context, and this conventional way of talking suggests that the particular action is a dependent variable, while the context is the "independent" or determining variable. But this view of how an action is related to its context is likely to distract the reader – as it has distracted me – from perceiving the ecology of ideas which together constitute the small subsystem which I call "context". … It is important to see the particular utterance or action as *part* of the ecological subsystem called context and not as the product or effect or what remains of the context after the piece we want to explain has been cut from it (1972, p.338).

This is how the world is made up, Bateson argues, a very complex network (not chain) of entities which have this sort of relationship to each other. Part and whole are different logical types, and any description of change – for example learning – will disclose a hierarchy of logical types. He takes the idea of logical typing from Russell and Whitehead. A class is a different logical type, at a higher level of abstraction than the members it classifies. We build this into a hierarchy of classes – classes of classes – classes of classes of classes etc. The more abstract class always classifies the less abstract one. Actions, items, persons, perceptions, etc., are organized by us into a pattern, a contextual structure, and it is this structure which informs us how to interpret an utterance. It is a joke? Was it meant to be compliment?

Bateson uses a variety of synonyms for context; he speaks of it as whole, system, form, pattern and relationship. The process of structuring a context is called punctuation. Punctuation, according to Bateson, is the basic epistemological act by which we draw distinctions and create differences and order. This is worth quoting at length:

> … we might ask: "What circumstances determine that a given scientist will punctuate the stream of events so as to conclude that all is predetermined, while another will see the stream of events as so regular as to be susceptible of control?" … "What circumstances promote that specific habitual phrasing of the universe which we call 'free will' and those which we call 'responsibility,' 'constructiveness,' 'energy', 'passivity', 'dominance', and all the rest? For all these abstract qualities, the essential stock-in-trade of the educators, can be seen as various habits of punctuating the stream of experience so that is takes on one or another sort of coherence and sense (1972, p.163).

When we punctuate, we turn streams of events into sequences that are often linear, with beginnings and endings. We see something as cause and another thing as effect, or as stimulus, response and reinforcement, thereby ordering them and making sense of what is happening. In human life there are signals whose major function is to classify contexts; i.e., provide us with information of which context we are in. These are called context markers. They tell us which context we are in and how to understand the utterance or action – a wink can inform me that I am to interpret an utterance as a joke, in another context I might have taken it as an insult. Human constantly meta-communicate, Bateson observes, to make sure we know which context we are in. Behavior can be organized in contexts we can call "guessing", "discrimination", "play," "exploration" and the like (1980, p.135). Problems arise if the context belies the message, for example a teacher who encourages students to be critical but at the same time structures the context such that the students who are critical are continually put in the wrong and their critical utterances disqualified. The context is "the name of the game",

it is what the situation is all about. And we rely on signals that might be extremely subtle, even invisible, to pick it up.

Consider Bateson's discussion of Pavlov's dogs (1980, p.133ff). How should we describe what the dog is doing and what is happening to it? We may say "the dog discriminates between ovals and circles" and at some point we may say with Pavlov that "the dog's discrimination breaks down". Here we have a case of a jump from a description of a particular incident (discriminates) to a generalization in the form of an abstraction (discrimination) that now seems to reside inside the dog. We have made a jump in logical type, from particular to general, member to class. What has happened? From the dog's point of view; first he learns to discriminate between O and C. The dog learns that this is a context for discrimination; that his task is to act on the difference between O and C and that success will be rewarded. But when there no longer is a perceptible difference between O and C, it is no longer that kind of context. The context has changed. And the dog fails to make the jump from "context for discrimination" to "context for guessing". This would have been a discontinuous jump from one kind of learning (discriminating between visual stimuli) to another kind of learning (learning the context). In all such cases, Bateson says,

> … the step from one logical type to the next higher is a step from information about an event to information about a class of events or from considering the class to considering the class of classes (1980, p.137).

In the Pavlovian case, the dog no longer knew what was expected of him, and clearly this was disturbing. The system – the name of the game had changed; by an external intervention we might say.

Learning character

The lesson of this very brief foray into system theory is that character is learned in the way we learn context, not in the way we learn actions, skills or knowledge. Character is a different logical type than actions.

Learning, Bateson insists, is *necessarily* hierarchical. That is because it contains components of trial and error, and error is expensive. In school errors can be socially expensive, say, if you fail an exam, do badly at tests or break the dress code. Second-order learning is therefore needed to reduce the amount of trial and error involved in achieving first-

order learning. For example, one might learn to find the correct answer to teacher-posed questions much faster.

Bateson posits five levels of learning, designated Learning 0, I, II, III and IV. This is not a stage theory, nor does it have an in-built logic of evaluation such that Learning 0 should be inferior to Learning III, for example. This is a hierarchical *logic*, applied to learning. Learning 0 is responding to stimuli but making no changes based on experience or perception. Highly skilled professionals who use pattern recognition developed over years of practice in response to events, exhibit Learning 0. Learning I denotes regular learning of knowledge and skills in school, such that when posed questions or given tasks you have new choices to make. Formally, Learning I is a change in specificity of response by correction of errors of choice within a set of alternatives. The set of alternatives itself does not change; you just pick a different alternative in a multiple choice test.

Learning II denotes changes in the process of Learning I; formally it is a corrective change in the set of alternatives from which choice is made, or a change in how the sequence of experience is punctuated. The phenomena of Learning II are of the greatest interest to us here. This is context learning; how we learn to expect our world to be structured in one way rather than another. It is a discontinuous form of learning, like the jump from a context of discrimination to a context of guessing – a new set of alternatives is identified, and within this set new choices of actions, perceptions and responses can be found. Punctuation is not true or false, Bateson says, and the contexts in question might be perceived differently by different people.

The phenomena of Learning II, Bateson says, are a major preoccupation of educators (or should be). And here we have finally come to *character*. By learning ways to punctuate sequences and by learning certain context markers, we learn what is expected of us and we adapt. We learn the principles, governing rules or the patterning of a context. Thus, Learning II will always be present during Learning I, as the context that classifies what is going on in Learning I. Descriptions of character structure "are derived not from what the subject has learned in the old simple sense of the word "learning," but from the *context* in which the simple learning has occurred" (2009, p.217). If we intervene in classrooms to bring about enhanced virtues and improvement of character, we treat those phenomena as if they are cases of Learning I. We have our students read selected texts, they discuss them, perhaps we do role-play, and finally the students are tested so we can evaluate the effectiveness of the intervention. This series of events defines a commonplace educational pattern: the teacher acts, the students respond, the teacher reinforces or sanctions the responses, and in the end the

students are tested for the effects on their individual virtues and character. This seems like a context of instrumental reward; I do what I am expected to do, and that brings a reward in the form of a good grade. With repeated contexts like this, I learn that the world is made up of contexts in which I can act instrumentally. I will mold contexts I meet to fit my expected punctuation, which will thus be reinforced. Learning II is self-validating, Bateson suggests, and therefore practically impossible to eradicate. It takes Learning III to do that; a questioning of the deep habits, premises and traits that are the results of Learning II.

Thus, we might find that intervening to enhance virtues and improve character might have adverse effects. These are not Learning I phenomena, they cannot be taught in the same way as we teach mathematics or languages. There is nothing wrong with having students read texts, participate in classroom discussions and do role-play, and they might well be able to tell you what they learned from it. But character is a different logical type. I have suggested it consists of an aggregate of virtues, features and traits that form an individual's nature or personality and this is learned or formed by picking up on subtle context markers to learn the name of the game. Actions might obey simple rules of reinforcement and/or sanction, but *ways of organizing* one's actions – behavior we might describe as 'discrimination,' 'guessing,' 'dependency' or 'play' – do not, nor are they testable by standardized tests, I surmise. If Learning I is structured as a context of instrumental reward, despite all good intentions behind the intervention, then that is how their character will be formed: by learning instrumentality. To get at people's character, we instead have to pay close attention to the context we structure. If we want to know *why* people think the way they do, we have to ask what contexts, what part-whole relationships, will promote that specific habitual perception, such that we come to endow some actors with responsibility and others with obedience. The answer to such why-questions lies in the spelling out of what relationships hold between the given part and the whole, what contexts of learning that might understandably inculcate the beliefs or traits in question. In principle the explanation of human behavior is always in terms of ever wider contexts and meta-contexts. It is not a matter of identifying causes.

The primacy of relationship

Things get even worse from an interventionist's point of view. If we think of character and virtues as traits belonging to an individual and try to measure them, we are guilty of a domitive mistake. The term is taken from Molière's play *Le Malade Imaginaire* (The Hypochondriac, 1673), where a doctoral candidate explains why opium causes sleep by

stating that is contains a dormitive principle (*virtus dormitiva*). Since Molière this has been the name for a kind of explanation that in reality explains nothing – opium causes sleep because it is its nature to make people fall asleep. For Bateson, we make a dormitive mistake when we explain people's behavior by referring to a character trait rather than by laying out the contingency patterns of their context. We have thereby shifted from the interpersonal field to some inner tendency, principle or instinct which is made "responsible" for the action, like "discrimination" being inside the dog and capable of breaking down. Such a way of thinking represents a reification of the inner tendency by treating it as a fixed entity with causal powers. Reification is for Bateson a big epistemological sin; to treat patterns, forms and differences as if they were things. Boys get into fights because they are aggressive; students disrupt the classroom because they are bored or unmotivated, a spouse nags because he or she is a nagger, etc. Such explanations, Bateson proclaims, are plain nonsense and only serve to hide the real questions (1980, p.146).

If we want to talk about character and personality traits, we must minimally talk about two persons and what happens between them. Descriptions of personality traits and virtues actually consist of descriptions of extracted halves of larger relationships. It makes no sense to think of somebody as being "friendly," "dependent," "proud," "courageous" or "virtuous" outside a relationship that involves at least two people. When we attribute personal properties to people, we have – more or less arbitrarily – cut out a chunk from a larger sequence of experiences and events and turned into a dormitive principle; reified it into a fixed inner tendency that can be used to explain behavior. We thus ignore the other half of the relationship, we get half the story, so to speak. What we have to do, if we want to understand what is going on in a relationship, is to describe the punctuation of all involved parties: "*All* characterological adjectives are to be reduced or expanded to derive their definitions from patterns of interchange; i.e. from combinations of double description," Bateson declares (1980, p.147). From a systemic point of view it makes no sense to measure individual character improvement – we will commit a dormitive mistake, and we will simply not get at what we want because we neglect the relationship.

It may well be that interventions in general thrive on dormitive mistakes. Most causal hypotheses are in fact dormitive mistakes, Bateson thinks. This is exacerbated by how we tend to think of interventions as two-unit affairs (the X-Y relation): we intervene, and the students respond. The teacher teaches, and the student learns. But the systemic principle of circuits of interaction tells us that the minimum unit of interaction at least consists of three elements, which Bateson calls stimulus, response and reinforcement. The second is the

reinforcement of the first, and the third is the reinforcement of the second. Response by a student hence reinforces the stimulus provided by the teacher (the intervention). This happens in all human interactions – together we create a pattern which classifies our messages in the interaction – and it may be that EBP is not able to accommodate this feature, especially not the EBP orthodoxy which demands fidelity of implementation (the intervention must be implemented according to predetermined procedures).

Conclusion

I have in this paper discussed whether it is possible to conduct interventions in order to enhance students' virtues and improve their character, and I have tried my best to problematize the idea. My arguments have followed two different paths. First, in the context of EBP, I have looked at the logic of intervention. Interventions are predicated on a number of more or less explicit causal presuppositions. Importantly, focusing on X means ignoring other factors in the pre-existing context, some of which may be necessary for X to reach Y at all. Interventions tend to reduce complex processes to a matter of input and output, because we ask *whether* they work and not the more fundamental *how*. I conclude that interventions generally are not impossible, but more demanding than we think.

Second, and more fundamentally, I have used system theory to argue that character formation (improvement) is not a kind of phenomenon that lends itself to interventions at all. Character is formed as a result of Learning II, context learning, not as a result of Learning I (knowledge and skills acquisition). It is a different logical type and cannot be taught and learned the same way as math or geography. From a systemic point of view it makes no sense to measure individual character improvement – character traits belong to relationships and if we focus on the individual we all we get is an extracted half of a process. In fact, from a systemic point of view the whole idea of interventions as two-unit affairs (intervention leading to student result) may be seen as an over-simplification. It neglects the circuits and feedback mechanisms of human interactions. Character formation rather requires careful attention to the contexts we create and what messages those contexts convey to students about what is valuable, appreciated and expected.

References

Ashby, W.R. (1964): *An introduction to cybernetics*. London: Methuen

Bateson, G. (1972): *Steps to an ecology of mind.* New York: Ballantine Books

Bateson, G. (1980): *Mind and nature. A necessary unity*. Glasgow: Fontana/Collins

Bateson, G. (2009): Conventions of communication: where validity depends upon belief. In J. Ruesch & G. Bateson, *Communication. The social matrix of psychiatry* (212-227). New Brunswick: Transaction Publishers

Bertalanffy, L. von (1984): *General system theory (revised edition)*. New York: George Baziller. First published 1964

Cartwright, N. (2001): *The Dappled World. A Study of the Boundaries of Science*. Cambridge: Cambridge University Press

Cartwright, N. (2009): How to do things with causes. *Proceedings and Addresses of the American Philosophical Association*, 83, 2, 5-22

Cartwright, N. (2012): Will this policy work for you? Predicting effectiveness better: How philosophy helps. *Philosophy of Science*, 79, 5, 973-989

Cartwright, N. & Hardie, J. (2012): *Evidence-based Policy. A Practical Guide to Doing it Better*. Oxford: Oxford University Press

Hitchcock, C. (2001): Review of Judea Pearl: Causality. Models, Reasoning and Inference. *The Philosophical Review*, 110, 4, 639-641

Hitchcock, C. (2007): Prevention, preemption, and the principle of sufficient reason. *The Philosophical Review*, 116, 4, 495-532

Pearl, J. (2001): Direct and indirect effects. *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence* (411-420). San Francisco: Morgan Kaufman

Pearl, J. (2009): *Causality. Models, reasoning and inference*. Cambridge: Cambridge University Press

Sloman, S. (2005): *Causal models. How people think about the world and its alternatives.* Oxford: Oxford University Press

Woodward, J. (2008). Causation and manipulability. In E. Zalta (ed), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/causation-mani/ Retrieved March 20, 2012