# Cheating and Neutralization

## Mathieu Bouville

### University of Cambridge

### <mathieu.bouville@gmail.com>

**Abstract:** The concept of neutralization —originally developed in the field of criminology— has been widely used in the context of student cheating. Cheaters are said to believe that cheating is wrong yet to deny that what they did is wrong or to blame someone else. I argue that authors who make claims of findings of neutralization generally cannot substantiate them. They over-interpret their data, in particular taking 'does not contradict the theory' as meaning 'proves the theory'. I also point out that purely empirical data cannot be enough, since neutralization requires some theory of what is wrong behaviour, and no such theory has been provided so far.                     © Mathieu Bouville, 2008

## INTRODUCTION: ON NEUTRALIZATION

The concept of neutralization has been widely used in the context of cheating for over two decades (Carpenter *et al.*, 2006; Davy *et al.*, 2007; Granitz and Loewy, 2007; Haines *et al.*, 1986; Jordan, 2001; McCabe, 1992; Murdock and Anderman, 2006; Pulvers and Diekhoff, 1999; Storch *et al.*, 2002; Whitley, 1998). This theory comes from the work of Gresham Sykes and David Matza (1957) in the field of criminology: "much delinquency is based on what is essentially an unrecognized extension of defenses to crimes, in the form of justifications for deviance that are seen as valid by the delinquent but not by the legal system or society at large" (the whole sentence is in italics in the original, p. 666). Neutralization is essentially a form of bad faith — what Stephens and Nicholson (in press) call "knowing the right, doing the wrong".

When one reads the article of Sykes and Matza, the overwhelming impression is that the authors made a number of observations regarding juvenile offenders and concluded that extant explanations of criminal behaviour were simply not supported by facts. They did not believe that "the world of the delinquent is the world of the law-abiding turned upside down and its norms constitute a countervailing force directed against the conforming social order" (p. 666), in particular because criminals revere "really honest" people and may for instance exhibit "a fierce attachment to a humble, pious mother or a forgiving, upright priest" (p. 665).

In the case of the literature on cheating the impression is not one of insight, but rather that the model precedes the facts: 'we know that our results will confirm this model' superseded 'let's find a model that explains our observations'. Saying that one's results show neutralization becomes a stylistic convention. For instance, Passow *et al.* (2006) wrote "This attitude, called *neutralization*, has been found to be an important influence on college students' cheating behavior. Our results also support this finding." (p. 677, references removed) but never mention

neutralization again in their article. It seems that the reason for using the concept of neutralization in the case of cheating is that cheating is seen as *a priori* roughly similar to crime rather than because cheaters actually exhibit the behaviours and attitudes found in criminals by Sykes and Matza (1957). The question of whether neutralization is a suitable model for student cheating is never really asked. Instead one assumes that it should be rather suitable and one moves on.

Given that many studies on cheating mention 'neutralization' (Carpenter *et al.*, 2006; Davy *et al.*, 2007; Granitz and Loewy, 2007; Haines *et al.*, 1986; Jordan, 2001; McCabe, 1992; Murdock and Anderman, 2006; Pulvers and Diekhoff, 1999; Storch *et al.*, 2002; Whitley, 1998), it is important to find out whether it is indeed a valid model of students' behaviour, i.e. whether those who claim that cheaters neutralize have the necessary evidence and valid inferences from them to demonstrate that neutralization really occurs with cheaters. Two questions arise. Is what researchers find actually neutralization (as described above) or is it something else? Can empirical works validly infer their conclusions (whatever they are) from their survey data? I will address these in turn.


## IS WHAT CHEATERS DO NEUTRALIZATION OR SOMETHING ELSE?

The first question to ask is whether the results obtained in the case of cheating really correspond to the neutralization theory of Sykes and Matza (1957) or rather to some other model.

### Do cheaters revere the really honest?

Sykes and Matza noted that "the juvenile delinquent frequently accords admiration and respect to law-abiding persons. The 'really honest' person is often revered, and […] un-questioned probity is likely to win his approval" (p. 665). Is this true of cheaters? do cheaters revere honest students? In the case of the cheater, would one say that "unquestioned probity is likely to win his approval" or his laughter? The fact that juvenile offenders revere honesty incited Sykes and Matza to look at the discrepancy between this belief and their acts — they found "difficulties in viewing delinquent behavior as springing from a set of deviant values and norms" (p. 664). But in the case of cheating, authors do not notice that cheaters revere honesty and then question the discrepancy: they hold that cheaters neutralize, and if this requires that cheaters admit that probity is superior then it will be assumed that they do.

For instance, Davy, Kincaid, Smith, and Trawick (2007) wrote that "Those who neutralize profess to support a societal norm but rationalize to permit them to violate that norm." (p. 285). However, they never check that cheaters in fact support the norm against cheating: nothing in their survey addresses this point. Therefore they cannot show that there is neutralization, according to their own (correct) understanding of the neutralization concept. Donald Carpenter and his coworkers (2006) found that "71.0 percent of students either agreed or strongly agreed with the statement, 'it is wrong to cheat no matter what the circumstances' ", from which they concluded that "approximately three out of ten students would rationalize situations in which cheating would be acceptable […]. This suggests that students are willing to engage in behavior that they believe to be wrong" (p. 187). In fact, this implies that 29% of university students do not agree that cheating is always wrong, so if these students cheat they do not "engage in behavior that they believe to be wrong": they explicitly answered that they did not hold such behaviour to be always wrong. Similarly, Haines *et al.* (1986) claim that "the use of such [neutralization] techniques conveys the message that students recognize and accept cheating as

an undesirable behavior". But to show that there is neutralization, they must first show that the students "accept cheating as an undesirable behavior". So taking neutralization to show this is circular. It is a wide-spread error in issues of cheating to assume both that cheating is obviously wrong and that everyone agrees with this view (Bouville1, Bouville2, Bouville3).

### Proving neutralization requires theory

Neutralization could be invoked in situations other than cheating: one may for instance frame self-defence as neutralization, saying something like 'some murderers shift blame to the victim by claiming that they had no choice but to murder the victim'. Evans and Craig (1990) found that teachers are less likely to hold themselves responsible for the cheating of their students than students are to blame the teachers — can one conclude that teachers use neutralization to shift blame to the students? Jerald Greenberg (1993) studied the case of students who were employed to do clerical work. The students were allowed to take their pay themselves. Some of them had their salary arbitrarily decreased compared to what was initially promised to them. Greenberg observed that "equitably paid subjects took precisely the amounts they were allowed to take, whereas underpaid subjects took more than they were permitted (i.e., they stole)" (p. 81). Students took was they deserved, in some cases this meant not stealing, in others it meant stealing. Most people would see this as justice, as righting a wrong. But this can as easily be framed as neutralization: thieves neutralize their crime by shifting blame to their victims. Very similarly, Robin Hood could be said to neutralize his wrongdoing by shifting blame to the wronged (the rich and the usurper).

Sykes and Matza recognized that certain explanations for one's deeds (such as self-defence) are not forms of rationalization. They draw a line between valid and invalid justifications: the latter are what they call neutralization. This sets a limit on what qualifies as neutralization (without it about anything would qualify, making the concept absurd). This is in sharp contrast with research on cheating. For instance, Davy *et al.* label 'neutralization' their survey questions regarding whether cheating may be acceptable in this or that situation, and they assume (but do not show) that all of these are illegitimate. Likewise, Table 3 in Carpenter *et al.* (2006) gives "percentage of students agreeing with neutralization statements", i.e. who disagree that cheating is wrong in a number of specific situations. In their Table 1, Carpenter *et al.* give the "percentage of students defining a behavior as 'Cheating' " and the "frequency of cheating behavior" for certain behaviours, i.e. they ask students whether a given behaviour is cheating and then label it 'cheating' in any case.

Jensen *et al.* (2002) quote a high school student: "I'm a dedicated student, but when my history teacher bombards me with 50 questions due tomorrow or when a teacher gives me a fill-in-the-blanks worksheet on a night when I have swim practice, church, aerobics —and other homework— I'm going to copy from a friend!" (p. 210). Similarly, Ashworth *et al.* (1997, p. 202) found that "cheating is taken to be excusable where units are seen to be of marginal importance, or badly taught". Whether these are valid reasons for cheating is never asked.

In fact, no empirical study can show certain deeds to be wrong, so that no empirical work can, on its own, distinguish between rationalization and valid justification. Some theory is needed to determine whether certain circumstances can actually justify cheating. In the case of crime, such theoretical work exists in ethics and in law, but so far nothing has been done concerning cheating (see Bouville1). It is not my purpose here to ask whether there actually are valid justifications for cheating: my point is that unless one proves that these do not exist one cannot know where to draw the line between valid justifications and rationalizations.

Consequently, one cannot show that neutralization occurs. Some theory is thus necessary — empirical studies are not sufficient.

### Proving neutralization requires specific facts

Storch, Storch, and Clark (2002) consider that students who cheat because "professors do not assign grades fairly" are trying to "shift attention from the deviant act to the motives and behaviors of those that disapprove of the violation". Storch *et al.* never notice that there is something rotten in a class where this actually happens, so that cheating may not be the main problem (also see Parameswaran, 2007). Researchers never investigate such claims either, thereby showing that they see actual circumstances as irrelevant: only what the students offer as explanation for their cheating matters, not facts.

In the case of a claim of self-defence, on the other hand, one must first determine whether self-defence is an acceptable reason for killing (theoretical question) and then whether there was indeed a threat on someone's life (empirical question). If the answers to both questions are yes then there cannot be neutralization because the killing was completely justified. Plainly, if neither question is asked then it is impossible to claim that there is neutralization. Comparing the specific facts to the theory describing circumstances that make killing forgivable is necessary to decide if this was murder or self-defence. There can be neutralization only in the case of a crime, not in the case of a non-criminal act such as self-defence.

By claiming that students who mention circumstances that supposedly 'justify' cheating necessarily try to rationalize their cheating one assumes that instructors are never wrong (never incompetent, never lazy, never unfair, etc.) and that only students can be blamed for cheating. This may be true but, since this is crucial to the claim of neutralization, it should be argued for explicitly (also see Bouville1, Bouville4).

## Proving neutralization

I showed that empirical studies of cheating are unable to establish instances of neutralization because they have neither the necessary theory nor specific enough factual information. The other question to ask is whether authors really obtain empirical evidence that can substantiate the claim of neutralization. In particular one must ask whether the data *prove* their conclusion or are merely compatible with it.

### Getting out what one puts in

Jeanette Davy and her coworkers (2007, p. 295) stated that "Prior Cheating has a significant positive effect on Neutralization (.792)". In their study, they determined whether students neutralized by asking them whether cheating is acceptable in a number of circumstances (see their Table 1). The positive correlation they found empirically is thus between cheating and belief that cheating is acceptable in some cases. In other words, they found that students who hold cheating as more acceptable cheat more, which is not a very surprising conclusion. The only reason why they instead claimed that "prior cheating was also positively related to neutralization, further supporting arguments that the more individuals engaged in unethical/ dishonest behaviors, the greater was their need to rationalize and justify those behaviors" (p. 297) is because they decided beforehand to interpret their results in terms of neutralization. As Austrian philosopher Karl Popper (1935) pointed out, "observations, and even more so observation statements and statements of experimental results, are always *interpretations* of the facts observed; that they are *interpretations in the light of theories*. This is one of the main reasons

why it is always deceptively easy to find *verifications* of a theory" (p. 90, emphasis original). Let us assume that I want to prove that A implies B and empirically find a correlation between A and B, I may wish to conclude that I proved that A indeed implies B. In fact, the data are just as compatible with 'B implies A' or the existence of some C that entails both A and B. This is essentially the drift of Davy and her coworkers: their empirical results were not incompatible with the theory they wished to prove so they concluded that the data supported the theory. If one were to write in plain English what they actually found (without the superfluous reference to neutralization), the outcome would be that students who take cheating to be more acceptable cheat more. Davy *et al.* get out a stronger conclusion simply because they put it in.

They also wrote (p. 286): "If one is not engaging in unethical behavior, there is no need to develop rationalizations to neutralize any sense of disapproval by oneself or others". If one removes the negative connotations of 'rationalizations' and 'neutralize', this gives: 'if one is not engaging in unethical behaviour, there is no need to provide justifications to counter any sense of disapproval by oneself or others'. But this statement is plainly false: for instance, someone killing in self-defence will have to provide justifications to avoid being sent to jail. Their statement holds only if disapproval is always justified, i.e. if what seems unethical always is. But this simply assumes that there can be only rationalizations and no valid justifications — they assume what they are supposed to prove. (One should also note that this is very similar to claiming that innocents do not need a lawyer, so someone with a lawyer is guilty. "Once we've decided that someone's action is morally wrong, her efforts to challenge that premise, no matter how well-reasoned, merely serve to confirm our view of her immorality" (Kohn, 2007) — very much in line with the Spanish inquisition school of epistemology.)

These examples show that the results are typically interpreted in the light of the theory one seeks to prove. Consequently they are (probably unconsciously) made to fit the theory. 'This piece of data can fit in the model if we interpret it this or that way' (i.e. it is not completely incompatible with the model) soon becomes 'this piece of data supports the model' — in actuality it is rather the model that supports the data.

### Definition of cheating

Carpenter *et al.* (2006) asked students what they considered cheating (their Table 1; also see Table 1 in Passow *et al.*, 2006) before asking them whether cheating is wrong under given circumstances (Table 3). "Copying an old term paper or lab-report from a previous year" (i.e. plagiarism) is a form of cheating for only 60.7% of the students. What then does it mean that "71.0 percent of students either agreed or strongly agreed with the statement, 'it is wrong to cheat no matter what the circumstances' " (p. 187)? It is possible that 39.3% of students do not hold plagiarism as cheating and an additional 29.0% consider it cheating but not wrong, i.e. that plagiarism not be wrong for 68.3% of the students. Or it may be that all those who do not say that cheating is wrong do not believe that plagiarism is cheating either; in this case 39.3% of students would say that plagiarism is not wrong. In other words, we do not know whether four or seven out of ten students consider that plagiarism is not wrong.

We do not know what students take to be acceptable since we do not know what they mean by 'cheating' when they say that cheating is wrong. Imagine that a burglar says that crime is always wrong but that what he does is not criminal (e.g. because he never hurt anyone): would he not be a good potential candidate for neutralization? Table 1 of Carpenter *et al.* seems to indicate that neutralization may take place in the definition of cheating itself. This is never taken into account.

### Neutralization or incoherence?

Carpenter *et al.* (2006) found that 30% of students strongly agree that "it is wrong to cheat no matter what the circumstances" yet only 23% strongly agree that cheating is wrong "even if the instructor has done an inadequate job of teaching the course" or "even if the instructor assigned too much material" (see their Table 3). A quarter of the students who hold cheating to be always wrong thus disagree that it is wrong in more than half the circumstances mentioned. These answers are not 'neutralization': they are incoherent. And incoherent answers are useless — nothing meaningful can be obtained from something meaningless—, so that a coherent answer must be obtained before anything is said on what the students think about cheating and in particular whether there is neutralization.

One may change the questions slightly and ask whether killing is wrong: people would answer positively. One would then ask if killing in self-defence is wrong and they would certainly answer negatively. If one points out the discrepancy people will probably say that killing is wrong in general even though there are special cases, such as self-defence. It is likely that when asked about killing or cheating out of the blue and in a very general way, people will give an unreflective answer (possibly the answer they think they are supposed to give). When asked about specific cases, they will be more likely to think about these situations in more precise terms and to ponder whether killing or cheating is in fact wrong in such cases. It is thus quite possible that when students say that cheating is wrong we should not interpret this as meaning that cheating is always wrong with no possible exception, but rather that cheating is generally wrong (like killing is generally wrong). Naturally, surveys are not interactive and cannot point out possible incompatibilities and ask the students what they really mean; so they cannot tell whether students neutralize or answer too fast — more in-depth methods (e.g. interviews) may be necessary.

### On real students

Jason Stephens and Heather Nicholson (in press) interviewed a student who "sees cheating as wrong but finds himself doing it and feeling guilty about it". The student said: "When I cheat, it's like I don't want to but then it still happens but then at the end I feel bad that I did it because I know that I'm not really learning anything". While this student clearly does something he takes to be wrong, he does not attempt to rationalize it: if he did, he would not feel bad about cheating. Another student is "simply not very interested in learning (or working hard at it) and he isn't much emotionally affected by his cheating, which he acknowledges is wrong". Both students cheat and both say that cheating is wrong. But neither seems to neutralize his cheating: one is too overwhelmed and the other is too underwhelmed. (Of course, Stephens and Nicholson also found students who seem better candidates for neutralization; but the fact that the neutralization theory does not apply to a large majority of students should give one pause.) Naturally, surveys lack the level of detail necessary to notice this. Furthermore, while their attitudes towards cheating are as different as they can get, these two students may answer questions on cheating in a similar manner and thus look the same in survey-based research. Perhaps, understanding cheating is like painting: you cannot do it by numbers.

## CONCLUSION: DO EMPIRICAL STUDIES SHOW NEUTRALIZATION?

The theory of neutralization of Sykes and Matza (1957) sprang from what they considered the failure of extant criminology theories — "The basic characteristic of the delinquent sub-culture, it is argued, is a system of values that represents an inversion of the values held by

respectable, law-abiding society" (p. 664). They instead believed that criminals revere "really honest" people. Is the same true of cheaters? In any case, no evidence has been provided to show that cheaters really admire the honest students. Sykes and Matza also admit that circumstances can really justify what may otherwise be a crime (e.g. killing someone in self-defence). Neutralization basically means pushing the separation between crime and self-defence to the point of making all crime justified. But this implies that in order to argue that someone is neutralizing one must be able to locate the border between crime and self-defence; otherwise one cannot claim that the cheater is trying to shift it. Some theory is needed before one can make claims regarding cheating and neutralization. There is also the issue of whether survey data are capable of *proving* the conclusion of neutralization. Oftentimes, the fact that empirical results are not incompatible with a theory is deemed sufficient proof that the data supported the theory; for instance, researchers conclude that cheaters need to make up justifications for their cheating, when they merely show that students who take cheating to be more acceptable cheat more. A particular issue with neutralization is that it is intrinsically paradoxical: it means doing something one considers one should not do. But this paradox can cross the line to literal absurdity: Carpenter *et al.* (2006) found that a quarter of the students who hold cheating to be always wrong disagree that it is wrong in many circumstances. One should also remark that surveys (the tool of choice for studying cheating) have intrinsic limitations. Jason Stephens and Heather Nicholson (in press) interviewed two students who cheat, but neither seems to rationalize his cheating. Moreover, even though their attitudes towards cheating are diametrically opposite, these students may answer survey questions in a similar fashion, i.e. look the same to the researcher.

Neutralization has become a stylistic convention rather than a genuine theory: its main characteristic is that it is something authors are supposed to mention in their paper. Consequently, authors find neutralization before they even have empirical results. Whether cheaters actually neutralize seems less important than boasting a finding of neutralization.

© Mathieu Bouville, 2008

## REFERENCES

Ashworth, P., Bannister, P., and Thorne, P. (1997) Guilty in whose eyes? University students' perceptions of cheating and plagiarism in academic work and assessment, *Studies in Higher Education* **22**, 187–203.

Bouville1: Why is cheating wrong? Available at http://arxiv.org/abs/0803.1530.

Bouville2 (2008) Plagiarism: Words and ideas, *Science and Engineering Ethics* **14**, 311–322.

Bouville3: Crime and punishment in scientific research. At: http://arxiv.org/abs/0803.4058.

Bouville4: Exam fairness. Available at http://arxiv.org/abs/0803.4235.

Carpenter, D. D., Harding, T. S., Finelli, C. J., Montgomery, S. M., and Passow, H. J. (2006) Engineering students' perceptions of and attitudes towards cheating, *Journal of Engineering Education* **95**, 181–194.

Davy, J. A., Kincaid, J. F., Smith, K. J., and Trawick, M. A. (2007) An examination of the role of attitudinal characteristics and motivation on the cheating behavior of business students, *Ethics and Behavior* **17**, 281–302.

Evans, E. D., and Craig, D. (1990) Teacher and student perceptions of academic cheating in middle and senior high schools, *Journal of Educational Research* **84**, 44–52.

Granitz, N. and Loewy, D. (2007) Applying ethical theories: Interpreting and responding to student plagiarism, *Journal of Business Ethics* **72**, 293–306.

Greenberg, J. (1993) Stealing in the name of justice, *Organizational Behavior and Human Decision Processes* **54**, 81–103.

Haines, V. J., Diekhoff, G. M., LaBeff, E. E., and Clark, R. E. (1986) College cheating: Immaturity, lack of commitment, and the neutralizing attitude, *Research in Higher Education* **25**, 342–354.

Jensen, L. A., Arnett, J. J., Feldman, S. S., and Cauffman, E. (2002) It's wrong, but everybody does it: Academic dishonesty among high school and college students, *Contemporary Educational Psychology* **27**, 209–228.

Jordan, A. E. (2001) College student cheating: the role of motivation, perceived norms, attitudes, and knowledge of institutional policy, *Ethics and Behavior* **11**, 233–247.

Kohn, A. (2007) Who's cheating whom?, *Phi Delta Kappan* **89**, 88–97.

McCabe, D. L. (1992) The influence of situational ethics on cheating among college students, *Sociological Inquiry* **62**, 365–374.

Murdock, T. B. and Anderman, E. M. (2006) Motivational perspectives on student cheating: Toward an integrated model of academic dishonesty, *Educational Psychologist* **41**, 129–145.

Parameswaran, A. (2007) Student dishonesty and faculty responsibility, *Teaching in Higher Education* **12**, 263–274.

Passow, H. J., Mayhew, M. J., Finelli, C. J., Harding, T. S., and Carpenter, D. D. (2006) Factors influencing engineering students' decisions to cheat by type of assessment, *Research in Higher Education* **47**, 643–684.

Popper, K. (1935/1959) *The Logic of Scientific Discovery*. Routledge.

Pulvers, K. and Diekhoff, G. M. (1999) The relationship between academic dishonesty and college classroom environment, *Research in Higher Education* **40**, 487–498.

Stephens, J. M. and Nicholson, H. (in press) Cases of incongruity: Exploring the divide between adolescents' beliefs and behaviors related to academic cheating, *Educational Studies*. DOI: 10.1080/03055690802257127.

Storch, J. B., Storch, E. A, and Clark, P. (2002) Academic dishonesty and neutralization theory: A comparison of intercollegiate athletes and nonathletes, *Journal of College Student Development* **43**, 921–930.

Sykes, G. M. and Matza, D. (1957) Techniques of neutralization: A theory of delinquency, *American Sociological Review* **22**, 664–670.

Whitley, B. E. (1998) Factors associated with cheating among college students: A review, *Research in Higher Edu cation* **39**, 235–274.