How Tobacco Health Warnings Can Foster Autonomy

Adrien Barton*, Division of Philosophy, Royal Institute of Technology (KTH) and Department of Philosophy and History of Science, Kyoto University

*Corresponding author: Adrien Barton, Division of Philosophy, Royal Institute of Technology, Teknikringen 78 B, S-100 44 Stockholm, Sweden. Tel.: +33 (0)7 61 27 30 77; Fax: +46 (0)8 790 95 17; Email: adrien.barton@gmail.com

I investigate whether tobacco health warnings' interference with autonomy is ethically justifiable in order to deter people from smoking. I dissociate first the informational role and the persuasive role of tobacco health warnings and show that both roles enable typical addicted smokers to better rule themselves, fostering their autonomy. The fact that some messages address people's non-deliberative faculties is therefore compensated by a larger positive influence on their autonomy. However, misleading messages are not ethically justified and should be avoided. Tobacco health warnings' effect on autonomy highlights an important difference between libertarian paternalism and classical paternalism.

Introduction

On 29 February 2012, a US federal judge ruled that regulations requiring large graphic health warnings on cigarette packaging violate free speech rights under the US Constitution. In his ruling, the judge argued that the government has other tools at its disposal to deter smoking, such as including simple factual information on the labels rather than gruesome images. Setting aside the legal issue and the problem of free speech for tobacco companies, this decision raises the question whether it is ethically acceptable to deter people from smoking with such warnings that seem to rely on irrational persuasion. This is an especially important question when considering that the World Health Organization recently identified health warnings on cigarette packages among the six key measures required to reduce smoking prevalence, while smoking remains the leading cause of preventable death in many developed countries.

I will analyze this new generation of tobacco health warnings in the context of a recent approach in political philosophy named 'libertarian paternalism' (Thaler and Sunstein, 2008). Libertarian paternalism proposes to affect people's choices by interventions interacting with their non-deliberative faculties and improving their welfare. Although Thaler and Sunstein have presented such measures as innocuous, it has been rightly pointed that they infringe on autonomy, and need therefore a careful justification on a case-by-case basis (see e.g. Hausman and Welch, 2010, or Grüne-Yanoff, 2012). I aim here at providing precisely such a justification for tobacco health warnings. The core of my argument will be that non-misleading health warning messages actually foster the autonomy of the typical addicted smokers, through both their informational and persuasive roles. This investigation will have significant consequences for the general study of libertarian paternalism, by showing that classical paternalism and libertarian paternalism differ in some important aspect.

This article is structured the following way. First, I will sketch a short history of health warning messages on tobacco products and their efficiency and identify the area of investigation. Then, I will present libertarian paternalism. Afterwards, I will detail the informational role of health warning messages, and then their persuasive role. The next section will be central in my argumentation, investigating the multiple ways how health warning messages interact with people's autonomy, and concluding that it results in a net gain of autonomy for the typical addicted smoker. I will then raise two possible objections against this argument and show that they are misguided. Finally, a last section will complement the argument by considering two marginal cases: the non-addicted smoker and the willing addict. The conclusion will recap the argument and expose the difference it implies between libertarian paternalism and classical paternalism.

doi:10.1093/phe/pht011
Advance Access publication on 17 April 2013

Health Warning Messages on Tobacco Products: A Short Presentation

Let us first sketch a brief history of health warning messages on cigarette packages. 1964 saw the release in the USA of the first report of the Surgeon General's Advisory Committee on Smoking and Health. Following that event, USA was the first nation to require a health warning on cigarette packages. The first warning was a small-sized text reading "Caution: Cigarette Smoking May be Hazardous to Your Health", and similar warnings were soon imposed in other countries and extended to smokeless tobacco products.

In December 2000, Canada became the first country to enforce graphic warnings on cigarette packaging. Nowadays, at least 40 nations, many in Europe, have imposed more eye-catching warnings than the original small-sized warnings, including graphic photos. In Australia, a law imposing so-called 'plain packaging' is implemented since December 2012: cigarettes are now sold in olive brown cartons with large graphic images covering 75 percent of the front of the pack and all of the back. Packs from different brands are only differentiated by the brand and product names, written in a standard format.

The efficiency of health warning messages on tobacco products has been largely investigated. To my knowledge, the most comprehensive study to date on this respect was made by Hammond (2011), who reviewed 94 articles on this topic. It concludes that prominent health warnings on the face of packages serve as a prominent source of health information for smokers and non-smokers, can increase health knowledge and perceptions of risk and can promote smoking cessation. Also, it states that comprehensive warnings are effective among youth and may help to prevent smoking initiation. Finally, it notices that pictorial health warnings that elicit strong emotional reactions are significantly more effective.[1] In 2011, Dr. Lawrence R. Deyton, Director of the FDA's Center for Tobacco Products, estimated (based on other countries' experience) that the new warning labels would prompt an additional 213,000 Americans to quit smoking in 2013, the (at-the-time) planned first full year with the graphic labels.

As one can see, the actuality of health warning messages on tobacco products is vibrant, and in need of ethical analysis that may complement and enlighten the legal and psychological investigations on the topic. This is precisely what this article will attempt to provide, by investigating how such warnings interfere with people's autonomy, and whether this is ethically acceptable or not. Cigarette being by far the most consumed tobacco product worldwide, I will concentrate on health warning messages on cigarette packs. In order to tackle this question, I will show first how imposing such warnings can be considered as a libertarian paternalistic intervention.

Health Warning Messages and Libertarian Paternalism

In order to introduce the topic of libertarian paternalism, one should first explain what a regular (that is, non-libertarian) paternalistic measure is. Elaborating on Dworkin (2010), we can say that an action toward an agent is paternalistic if it fulfills different conditions: first, it interferes with the liberty of the agent (this also includes changing the financial incentives in acquiring a product); second, it is done without his consent and third, it will improve his welfare. For example, some classical measures aiming at decreasing tobacco consumption—like imposing a special tax on tobacco goods—are instances of paternalism: they change the financial incentives in buying tobacco, they are done without the consent of the smokers and they aim (amongst other goals) at improving their health.

It should be noted that attempts to reduce people's smoking are not univocally paternalistic: they can be justified, for example, in order to reduce second-hand smoke absorption. Moreover, paternalism is not always problematic, especially when it concerns minors rather than adults. In order to strengthen my point, I will therefore focus in this article on the hardest case, namely justifying attempts to deter *adults* (rather than minors) from smoking through health warnings, *for their own good* (rather than in order to protect other people from second-hand smoke).

I will now show that health labels belong to a specific form of paternalism, namely libertarian paternalism (Thaler and Sunstein, 2008). For this, let us first notice the obvious fact that decisions are never taken by humans in a vacuum, but in an environment arranged in a certain way. The insight behind libertarian paternalism is that environments could be arranged in different ways so that people make spontaneously decisions which are better for themselves. For example, instead of arranging the pizzas and salads side-by-side in a cafeteria, the salads could be placed more prominently with the pizzas a bit further back. Such a change in the

location of food items can increase or decrease the consumption of a specific food by 25% (Thaler and Sunstein, 2008). The environment within which a choice is made is named 'choice architecture'. A 'nudge' is an aspect of the choice architecture that alters people's behavior in a predictable way, without forbidding or significantly changing their economic incentives (that is, the intervention must be easy and cheap to avoid). To my knowledge, nudges have not been precisely defined yet in a non-controversial way, but at least a significant subclass of them[2] consist of interventions that influence people by interacting with their non-deliberative faculties (see Grüne-Yanoff, 2012, for a related account): simply bringing new information to people can alter their choices when being processed by their deliberative faculties, but does not count as a nudge.

There are many ways of defining 'deliberative' and 'non-deliberative' faculties in the cognitive psychology literature. What we will say will fit with several of these accounts, but we will endorse here more specifically the one proposed by Kruglanski and Gigerenzer (2011), which characterizes deliberative faculties as requiring cognitive effort and being accessible to conscious awareness. Non-deliberative faculties, on the other hand, require relatively little cognitive effort and are not accessible to awareness. In particular, as emphasized by Kruglanski and Gigerenzer (2011), deliberative faculties should not be confused with rule-based faculties (as intuitive faculties may be rule-based too), nor with 'rational' faculties (as it may be rational to use non-deliberative faculties). Here, 'faculties' should be understood as 'faculties of judgment and decision' (excluding e.g. perceptual faculties).

Let us consider again the above example of arranging the food in a cafeteria; in such a scenario, people are nudged to eat more salads and less pizzas, as this choice is not taken deliberately. Libertarian paternalism can be defined in the following way: an intervention counts as libertarian paternalistic if it nudges people to make choices that improve their welfare.

By definition, classical paternalistic measures interfere with an agent's liberty. On the contrary, libertarian paternalistic measures are not liberty-reducing[3]: they keep all choice alternatives open. That said, libertarian paternalistic measures do not come without a price. By influencing people while addressing their non-deliberative faculties, nudges may interfere with their autonomy. Therefore, libertarian paternalism involves a trade-off between autonomy and well-being.

Before going further, this concept of autonomy needs to be defined a bit further. The general idea behind autonomy is the 'capacity to live one's life according to reasons and motives that are taken as one's own and not the product of manipulative or distorting external forces' (Christman, 2011). Drawing on this account, we can dissociate two components of autonomy. First, 'the independence of one's deliberation and choice from manipulation by others' (Christman, 2011), which I will call the 'independence' component; the word 'manipulation' being negatively connoted, let us use instead the more neutral word 'persuasion', which we will define as an intervention addressing non-deliberative faculties.[4] The independence component of autonomy implies that the judgments and decisions of a perfectly autonomous agent should not be influenced in a way that addresses his non-deliberative faculties. The second component of autonomy is the 'capacity to rule oneself' (Christman, 2011) in order to aim at some goals, which I will call the 'self-ruling' component. This capacity requires in particular to be well-informed, so that following a self-chosen rule will have the desired effects. In the rest of the article, self-ruling should therefore be understood as 'informed self-ruling'.

With these distinctions in mind, we can claim that libertarian paternalistic measures that address people's non-deliberative faculties interfere with people's independence component of autonomy, and therefore raise ethical worries. Imposing health labels on tobacco products amounts to arranging the choice architecture in which someone will take the decision to smoke or not: as a matter of fact, in most cases, someone who is about to smoke will see the warning. If one can show that these labels address people's non-deliberative faculties, this will prove that they are a libertarian paternalistic intervention. I will show that this is indeed the case, by detailing the informational role and the persuasive role of these messages.

The Informational Role of Tobacco Health Warnings

In many respects, tobacco health warnings can be described as a kind of 'anti-advertising' against tobacco products. As a matter of fact, like advertising, tobacco health warnings have two functions: informing and persuading (Chapman, 1996). On one hand, they aim at informing the consumer about the risks raised by tobacco products; and on the other hand, they aim at persuading him not to buy the product. Whereas their informational role appeals to people's deliberative

faculties, their persuasive role addresses their non-deliberative faculties. Let us start by detailing this informational role of health tobacco warnings.

A study led by the International Tobacco Control Policy Evaluation Project in 2009 showed that health warnings on cigarette packages were the second source of information after TV about the risks of smoking in a majority of countries. Moreover, evidence suggests that countries with pictorial warnings demonstrate fewer disparities in health knowledge across educational levels (Hammond, 2011); therefore, pictorial warnings appear to be an efficient tool in order to spread information about tobacco risks in a way that is socially just.

Tobacco companies have long considered that the initial small-letter health warnings were sufficient to inform the consumers about the hazards of smoking. However, it has been shown that large text-based warnings are associated with increased health knowledge (Hammond, 2011). This provides an important justification for the transition from small-letter warnings to large-letter ones that has been operated in many countries. Other stylistic aspects may be important too in order to fulfill this informational role. For example, contrasted colors, like black letters on a white background, have been found to increase comprehension (Hammond, 2011). To sum up, there is an important ethical justification for writing objective informational messages and also for writing them in large contrasted letters rather than in small characters.

Some other messages fulfill an important information function—although I will show later that they also raise a few ethical worries. Messages mentioning specific diseases that may be caused by tobacco belong to this category. As a matter of fact, Chapman and Lieberman (2005) noticed that it is not enough, for being adequately informed, to know that smoking increases health risks: one should also be aware of the specific diseases caused by smoking. It appears that smokers are poorly informed in this respect, even for the most well-known tobacco-related diseases. Therefore, messages like 'smoking causes gangrene' in Singapore (as well as images showing the possible extent of the diseases) fulfill an informational role. One could object that such a disease is rare (more will be said about that in the next section); however, it might be rational to decide to smoke or not to smoke on the basis of this specific information only. For example, someone could follow a simple strategy which is to decide on the basis of the worst possible outcome; it would be justified for him to decide to stop smoking when learning that smoking can cause gruesome gangrene.

The Persuasive Role of Tobacco Health Warnings

Now that the informational role of tobacco health warnings has been presented, let us turn to their persuasive role. I will not try to make an exhaustive investigation of all existing warning messages, but I will present here three possible ways for tobacco health warning messages to address non-deliberative faculties: using social mechanisms like submission to authority; using emotional persuasion and exploiting cognitive heuristics.

Authoritative Messages and Recommendation Power

A first step away from the informational role is made when authoritative commands appear on warnings, like 'Smoking is highly addictive, don't start', which exploit people's natural submission to authority. Authoritative message can take a more hidden form, related to what Grüne-Yanoff (unpublished data) has called 'recommendation power' (in the context of default choices). In a nutshell, the idea is that messages can give implicit signals about what the best action is. A good example would be the warning 'Your doctor or pharmacist can help you quit smoking', which—despite not being commanding—suggests to smokers that they should quit. People reading this message may be inclined to follow the recommendation without deliberating about it. But recommendations can be even less direct. For example, it has been shown that viewers often equate the size of the warning with the magnitude of the risk (Hammond, 2011). However, the notion of 'magnitude' of a risk is—at least to some extent—subjective, since it depends on people's valuation of the bad consequences. Moreover, people generally evaluate risk and benefits not as two different entities but as one general compound: when one says that something is risky, it implies not only that he thinks it has some probable bad consequences but also that he thinks the benefits of this activity are not worth these bad consequences (Gregory and Mendelsohn, 1993). Therefore, a large-sized warning message may communicate implicitly that smoking is not worth its benefits. But people may have different valuations of smoking's bad consequences and benefits: some people may fear premature death more than others, some people may take a higher pleasure from smoking than others, etc. Therefore, when reading such messages, people may be pushed to take a decision independently of their values, instead of deliberating whether this is a good decision given their own values.

More simply, the fact that an originally informational message is read over and over confers upon it a strong recommendation power and makes it persuasive rather than simply informational. Being told only once that tobacco damages health is informational; being told this message 2000 times is also persuasive.

To sum up, some health labels influence people's valuations in ways that address their non-deliberative faculties. This is an instance of a general problem for libertarian paternalism identified by Rizzo and Whitman (2008): choice architects are privileged in selecting which values and preferences are promoted by the nudges they design.

Messages Using Emotional Persuasion

Some messages rely on emotional persuasion. Psychological research has shown that graphic depictions of disease appear to be the most reliable way to elicit negative emotional reactions (like fear or disgust) to health warnings, which in turn have been associated with engaging in cessation behavior[5] (Hammond, 2011). However, this cognitive process addresses non-deliberative faculties, playing on emotional reactions rather than on thoughtful deliberations.

Purely textual messages can also rely on emotional persuasion. For example, a European tobacco warning message states: 'Cigarette smoke contains benzene, nitrosamines, formaldehyde and hydrogen cyanide.' It is likely that most readers have no idea of the composition or effects of such products. But since these names evoke dangerous chemicals, people will infer (rightly) that they are dangerous and react with fear or disgust. However, the right inference is taken by non-deliberative means that may lead to false beliefs in other contexts. People without any biochemical knowledge would also presumably react with fear if a warning would inform them (falsely) that cigarettes contain L-ascorbic acid (which is the biochemical name for vitamin C): in a given context, any specific chemical name can raise fear.

Messages Exploiting Cognitive Heuristics

Other warnings rely on mechanisms that are less social or emotional, but more cognitive in nature. It has been largely documented that in situations of uncertainty, people tend to rely on cognitive shortcuts named 'heuristics', instead of using fully their deliberative faculties. When used in a context that is not appropriate, these heuristics can lead to cognitive biases—that is, deviation from the normatively correct judgments. One common heuristic on which people rely is the 'availability heuristic' (introduced by Tversky and Kahneman, 1973), which uses the ease with which examples come to mind in order to make judgments about the probability of events. For example, someone may consider that crime is frequent in his city because he has heard several times recently the TV news mentioning crimes committed there. Although this availability heuristic may sometimes be beneficial (see e.g. the work on the fluency heuristic, cf. Schooler and Hertwig, 2005), the frequencies that events come to mind are usually not accurate reflections of their actual probabilities in reality.

I have shown above that health warning messages mentioning specific diseases caused by tobacco fulfill an important informational role; however, they may also lead people to overestimate the probability of these diseases when using the availability heuristics. This could be especially worrisome when health warnings mention some rare diseases, for example the above-mentioned Singaporean warning stating 'Smoking causes gangrene'. The incidence of the corresponding disease (named Brueger's disease) is relatively low (8 to 12 per 100,000 adults in the USA, most of them due to tobacco products). Such a risk communication exploits the same kind of bias as the many advertisements for lotteries that insist on the size of the main prize, without mentioning the low probability of winning it[6]; in this respect, it is misleading.

Even messages giving correct statistics may be misleading. A warning in Canada states '85% of lung cancers are caused by smoking'. Even if this figure is accurate, many people are likely to commit what is called the 'inverse fallacy' (Villejoubert and Mandel, 2002) and confuse it with the incorrect statement '85% of smokers will get lung cancers' (the inverse fallacy consists in confusing the probability of A given B with the probability of B given A). Before putting the blame on people's poor understanding of probabilities, it must be noticed that this warning message has little relevance to inform a smoker about his risks: if lung cancer had a very low prevalence, 85% of lung cancers could be caused by smoking while the risk of getting lung cancer from smoking would be very low (as it happens, around 13% of smokers will get lung cancer—an already alarmingly high statistics). Therefore, this message invites to commit the 'inverse fallacy', since the incorrect statement '85% of smokers will get lung cancers' would be much more relevant in this context (see Sperber and Wilson, 1995, for an account of how people try to maximize relevance in communication). To sum up, although it gives the correct numbers, this warning can be seen as a misleading statistics. It is not unlike a

former advertising for a French lottery that stated '100% of all winners have tried their luck'.[7]

Persuasive Role: Summary

In summary, several text warning messages rely on strategies that address non-deliberative faculties: this shows that they infringe on the independence component of autonomy, by interfering with the choice of smoking in a persuasive way; and therefore, that they can be counted as a libertarian paternalistic measure. However, it is important to dissociate two families of messages. Some of them may trigger *false* inferences by non-deliberative means and are therefore both persuasive and misleading; for example, 'Smoking causes gangrene' may trigger the false inference that it is likely that a smoker will get gangrene. Some others may trigger *true* inferences by non-deliberative means, and are therefore persuasive but not misleading; for example, the warning 'Cigarette smoke contains benzene, nitrosamines, formaldehyde and hydrogen cyanide.' may trigger the true inference that cigarettes contain dangerous chemicals. The first family of message seems more problematic, and indeed, I will show in the next section that they are. Now that these informative and persuasive roles have been carefully dissociated, I can analyze the multiple ways by which health warnings may interfere with people's autonomy and how problematic this is.

The Interference on Autonomy

In general, libertarian paternalistic interventions on an agent trade a partial autonomy loss for a gain in welfare. It may be difficult—although not necessarily impossible—to evaluate the ethicality of such interventions, as these goods (autonomy and welfare) are heterogeneous categories. It may well be that the exceptional burden of disease associated with cigarette consumption would justify some loss of autonomy (it may well even justify fully paternalistic measures like taxes on tobacco products). However, it would be more satisfying if such measures could be justified without putting in the same balance goods that belong to heterogeneous categories. This is precisely such a justification that I will attempt to provide here: I will show that although health warning messages infringe on the independence component of addicted smoker's autonomy, this is compensated by an increase in their self-ruling component of autonomy, leading to a net gain in autonomy.

For this, the interference on autonomy occasioned by health warnings needs to be investigated closely. First, I will show that the informational role of health warnings fosters people's self-ruling component of autonomy, whereas misleading health warnings infringe on this component. Second, I will argue that the persuasive role of non-misleading health warnings also fosters the self-ruling component of autonomy of the addicted smoker. Third, I will show that although health warnings infringe on the independence component of autonomy, they protect against another cause of infringement, namely the implicit advertisement in a cigarette pack's design. Fourth, I will claim that health warning messages are fully transparent, and that everyone can easily avoid them substantially.

How Health Warnings' Informational Role Fosters People's Self-Ruling Component of Autonomy

Mill (1859) already argued that 'labelling [a] drug with some word expressive of its dangerous character, may be enforced without violation of liberty', since presumably 'the buyer cannot wish not to know that the thing he possesses has poisonous qualities'. This illustrates that the informational role of tobacco health warnings fosters people's self-ruling component of autonomy: indeed, by knowing more about the risks of tobacco, a rational agent should be able to lead his life in a better-informed way (as a reminder, being well-informed is an important dimension of the self-ruling component of autonomy). Moreover, the informational role does not address people's non-deliberative faculties, and therefore does not infringe on the independence component of autonomy.

For symmetrical reasons, insofar as misleading tobacco health warnings trigger false inferences, they make people less well informed, and therefore decrease their self-ruling component of autonomy. Therefore, misleading tobacco health warnings should be limited as much as possible (see the Conclusion section for a further discussion of this point).

How Health Warnings' Persuasive Role Fosters Addicted Smokers' Self-Ruling Component of Autonomy

Let us now turn to the persuasive role of health warnings and show that it also fosters the self-ruling component of autonomy for a typical addicted smoker[8] (when it is not misleading). For this, we need to investigate more precisely how health warnings may influence smokers' actions or desires.

Drawing on Frankfurt (1971), Goodin (1989) established a distinction between smokers' desires: they have a compulsive desire to smoke (a 'first-order' desire), but they also have a desire not to desire to smoke (since this is a desire about desires, it is called a 'second-order' desire). That is, most smokers suffer from a discrepancy between their first-order desires and their second-order desires. The typical smoker may also have a first-order desire not to smoke (or to smoke less), but tobacco addiction being strong, his first-order desire not to smoke is weaker that his first-order desire to smoke. Therefore, he smokes.

The persuasive role of health warning messages addresses people's non-deliberative faculties. It will reinforce their first-order desire not to smoke (or to smoke less) relatively directly. For example, a disgusting image of a cancerous lung will generally reinforce the first-order desire not to smoke without any need of deliberation. In some cases, this will enable this desire to be stronger than the desire to smoke, and consequently, the smoker will stop smoking (or will reduce his smoking). The self-ruling component of autonomy of the addicted smoker would then be fostered, as he now acts according to a rule (given by his second-order desire) that he has set for himself.

How Health Warnings Protect Against Another Cause of Infringement on Autonomy

I have established that the persuasive role of tobacco health warning messages addresses people's non-deliberative faculties, and thereby infringe on their independence component of autonomy. However, I will argue here that this loss is not as important as one could think.

Large and graphic health warning messages hide parts of the classical packaging that is designed by tobacco companies to attract smokers; this is of course especially true for the plain tobacco packaging format that is implemented in Australia since December 2012, which prohibit any companies' own design. As revealed in industry documents, 'the tobacco industry fully appreciates that packs are the premier site for communicating with smokers' (Chapman and Lieberman, 2005): packs' design acts as an implicit advertisement. Insofar as health warnings cover parts of the packs, they diminish the non-deliberative influence of the classical packaging's design. Therefore, they replace one source of interference with the independence component of autonomy (the tobacco company's implicit advertisement in their packaging's design) by another interference with this component (the health warning message). Arguably, the interference caused by some of the health warning messages may be stronger than the one caused by the packaging's design; but it can still be concluded that the independence component of autonomy is not decreased as much as one could think by the introduction of large health warnings.

One can even notice that some packaging formats may be considered as *misleading* advertisements. In particular, the design of some 'light' cigarette packs (with e.g. light colors predominating) may suggest the false idea that they are not harmful for smoking. Insofar as health warnings cover parts of the packs, they protect against this misleading implicit message. Therefore, they foster the self-ruling component of autonomy of the smoker.

Transparency and Avoidability

Let us now turn to another worry related to autonomy. Hausman and Welch (2010) have pointed that the government should inform people of efforts to shape their choices when they are engaging into libertarian paternalistic measures: people have a right to know that they are under influence. Bovens (2009) introduced a further distinction: government should not only inform people about how it will try to interfere with their agency—this is called 'type transparency'—but also inform people every time it tries to interfere with their agency—this is called 'token transparency'. As a matter of fact, people should not only know that they are under influence but they may also have a right to know *when* they are being influenced. Moreover, token transparency enables people who do not appreciate this type of manipulation to avoid nudges. Knowing when he is under influence is important for an agent to direct his life, and therefore to protect his self-ruling component of autonomy.

Coming back to health warning messages on tobacco products, it should be noted that they fully satisfy both requirements of type transparency and token transparency. As a matter of fact, people not only know that the government is trying to reduce smoking prevalence through warning messages in general: they also know, every time they see a warning message, that the government is trying to reduce smoking prevalence through this specific message. However, a superficial analysis could suggest that health warning messages are very difficult to avoid despite their transparency: indeed, they act as a kind of 'portative anti-advertising', following the smoker wherever he brings his cigarette pack. But it can be objected that special cases that hide the warning messages of cigarette packs are commercialized in many

countries.[9] Therefore, a smoker can avoid these warnings substantially[10] if they wish to, for an insignificant price.

In conclusion, warning messages on tobacco products are both type and token transparent, and they can be avoided substantially at a negligible cost. Therefore, they raise little worries in this respect.

Interference on Autonomy: Summary

I have shown that the informational role of health warning message protects people's self-ruling component of autonomy; this role is therefore uncontroversial. On the opposite, misleading messages are ethically not justified: the government should not communicate messages that are clearly known to trigger false inferences.

However, the persuasive role of some non-misleading messages has a more complex interaction with autonomy. On one hand, they increase typical addicted smokers' self-ruling component of autonomy (when they are not misleading); and on the other hand, they interfere with people's independence component of autonomy, by addressing non-deliberative faculties (although not as much as one could think, as they also protect against the infringement on autonomy due to the implicit advertisement in the design of the cigarette packs). In order to determine their net consequences on autonomy for the typical addicted smoker, we need to put in a balance their effects on his independence and self-ruling components of autonomy. Arguably, the second component is more important for the addicted smoker. Indeed, the independence component of autonomy is valuable mainly[11] to the extent that it protects people from interference on their authentic choices—the ones they can recognize as expressing their own selves.[12] However, the first-order desire to smoke of an addicted smoker is the effect of an unwanted addiction and does not push towards an authentic choice. Therefore, it is arguably more authentic for an addicted smoker to lead his life following his second-order desire to desire not to smoke,[13] rather than to keep his first-order desires (and the actions to which they lead) safe from any influence addressing his non-deliberative faculties. One can consequently argue that the interference with his independence component of autonomy is compensated by the increase in his self-ruling component. All in all, the persuasive role of non-misleading health warning messages would then end up fostering the autonomy of a typical addicted smoker.

Apparent Asymmetries between the Autonomy Loss Caused by Health Warning Messages and Tobacco Addiction

I will now turn to two apparent dissimilarities between the respective actions of health warning messages and tobacco addiction, which could be used to counter the conclusion that the persuasive role of health warning messages end up fostering the addicted smoker's autonomy. I will rebut these two arguments.

Tobacco Addiction and Judgments

It looks like although tobacco causes a craving in the smoker (that is, a very strong first-order desire to smoke), it does not influence his judgments: an addicted smoker seems to be free to think by himself. Indeed, many smokers know that they are addicted and recognize that smoking endangers their health. On the contrary, it could seem that health labels exercise their influence in a more insidious way: they modify the judgment of people, changing their perception of tobacco risks, by addressing their non-deliberative faculties. I will show here that the situation is not as asymmetric as one could think: tobacco also modify a smoker's judgments by addressing his non-deliberative faculties, although in a quite subtle way.

To establish this, some findings of the research literature on smokers' risk perception need to be reminded. Its results could appear paradoxical: on one hand, several studies seem to show that smokers overestimate the numerical risk of smoking (Marsh and Matheson 1983, Viscusi 1990, McCoy *et al.* 1992); but on the other hand, smokers still resist the idea that these risks are personally relevant (Chapman and Liberman, 2005). For example, smokers are prone to believe that they have a lower risk of developing a smoking-related disease than the average smoker (Hansen and Malotte, 1986; McKenna *et al.*, 1993; Weinstein *et al.*, 2005). Also, they overestimate their life expectancy (Schoenbaum, 1997).

These apparently paradoxical findings are generally explained by the cognitive dissonance theory. On one hand, smokers know that smoking is dangerous (and people being notoriously bad with numerical probabilities, they overestimate the numerical risk of smoking); on the other hand, they are addicted and cannot stop smoking. They can solve the conflict between their action of smoking and their belief that smoking is dangerous by holding the additional belief that, although

smoking is dangerous in general, it is not too dangerous for *them*. For example, many smokers hold the false beliefs that exercise undoes most smoking effects, or that they are protected by some genetic factors (see Peretti-Wattel *et al.*, 2007). In a nutshell, smokers believe the practice is not too dangerous for them because they smoke, rather than smoke because they believe it to be not too dangerous for them (Pears, 1984).

Therefore, it is not true that addiction to tobacco only influences desires: by making people addicted, it also influences their judgments in a non-deliberative way and causes false beliefs. Both tobacco addiction and health labels influence judgments by addressing non-deliberative faculties.

Tobacco Addiction and Agency

Let us now turn to a second possible objection. The addiction caused by tobacco seems to be due to the natural biochemical effect of the plant, not to anyone's agency. On the other hand, health warning messages are caused by thinking agents—namely public health officials, who therefore interfere with smokers' agency. Isn't an influence due to a thinking agent ethically more problematic, in some respect, than a natural biochemical influence?

However, this view overlooks a well-known fact: employees of tobacco companies have intentionally manipulated the tobacco blend to enhance the effects of nicotine in cigarettes, thereby increasing the addiction of the smoker. Addiction to tobacco is therefore partially caused by an agentive intervention.

In summary, both addiction to cigarettes and health warning messages are (to some extent) agentive interferences and interferences with people's desires and judgments. The argument claiming for an asymmetry in these respects is therefore rebutted.

Marginal Cases: The Non-Addicted Smoker and the Willing Addict

The focus was until here on the case of the addicted smoker, by far the most frequent. However, in order to complete this account, two further special cases need to be considered. The first one is the case of the non-addicted smoker. The second one is the case of someone who actually *wants* to be addicted to tobacco—for example, someone who wants to follow a classical 'rockstar' lifestyle, which implies to live in the present, to not care about the future and to indulge in all the pleasures associated with this lifestyle. Such a person may want to smoke and even wants to get addicted to tobacco, since this is part of this very lifestyle.

When they are persuasive rather than informational, health warnings interfere with the independence component of autonomy of the non-addicted smoker or the willing addict; but they also infringe on their self-ruling component of autonomy, by influencing them not to smoke, despite the fact that the non-addicted smoker freely choses to smoke, and that the 'aspiring rockstar' wants to get addicted. In these both (marginal) cases, the autonomy of an agent is violated; one could therefore argue that from a strict deontological point of view, this is not acceptable.

There are two ways to answer this objection. The first one would be to adopt a moderate consequentialist point of view, and thereby justify the violation of autonomy of this small minority by an appeal to the autonomy-fostering of the greater majority of addicted smokers. Although autonomy is a central value in deontological ethical reasoning, it can also be considered as a good in consequentialist ethical reasoning. Indeed, John Stuart Mill claimed that autonomy is 'one of the elements of well-being' (cf. Christman, 2011). It is therefore not absurd to consider that a small loss of autonomy of a small minority can be compensated by a greater gain of autonomy for the overwhelming majority. The second way to answer this objection is to notice that the autonomy loss for non-addicted smokers and willing addicts can be substantially avoided, since, as it was noticed, health warnings can be easily and cheaply avoided by placing the cigarette pack or the cigarettes in another case designed to this effect.

Conclusion

The original question 'Is tobacco health warnings' interference with people's autonomy ethically justifiable in order to deter them from smoking?' can now be answered. I have first highlighted that convincing minors not to smoke, or convincing adults not to smoke with the intention of protecting their surroundings from second-hand smoke, was ethically less problematic than using health warnings in order to deter *adults* from smoking, *for their own good*. Therefore, I have focused on this harder case. I have argued that warning messages foster people's self-ruling component of autonomy in two ways: first, by fulfilling an important informational role; and second, when they are not misleading, by influencing in a non-deliberative way typical smokers' actions so that they fit with their second-order

desires to desire not to smoke. Admittedly, these messages infringe on people's independence component of autonomy—but not as much as one could think, as they protect against another attack on autonomy, namely the implicit advertisement in the cigarette's packaging design. Moreover, the self-ruling component of autonomy of an addicted smoker is more important than his independence component of autonomy; therefore, health warning messages result in a net gain of autonomy for the typical addicted smoker. Thus, an ethical justification of health warnings has been provided—not only for the original small ones but also for the large-print and graphic ones, and even for the most extreme format, namely the 'plain packaging' that is implemented in Australia since December 2012.

I have also emphasized that warning messages should be designed to be as little misleading as possible—that is, they should trigger as few false inferences as possible. Some of the existing messages should therefore be corrected. It would not only be unethical to use plainly misleading statements: it could also be perceived as some kind of propaganda, which may decrease these warnings' efficiency or lead to a loss of trust in the state on the long-run.

Admittedly, when considering the ethicality of tobacco health warnings, one should also deal with other considerations than the infringement on people's autonomy. For example, health warnings—especially pictorial ones—may be very unpleasant to look at; however, this loss in well-being is arguably largely compensated by the gain in lives saved through health warning messages. Also, health warnings may contribute to the stigmatization of smokers; however, it might be argued that this compensates for long years of tobacco advertising that have tried to present the smoking lifestyle on a positive light (see Voigt, forthcoming, for a more extensive treatment of this question).

It should also be understood that the interference of health warning messages with autonomy are justified because of the specificities of tobacco products; therefore, health warning messages could not be imposed on other products without a strong, independent justification. For example, the fact that labeling tobacco products is ethically permissible does not imply that e.g. alcohol, junk food or saccharin labeling is also ethically permissible.

More generally, this article points to an important difference between regular paternalism and libertarian paternalism. Grüne-Yanoff (2012) has argued that libertarian paternalism is similar to regular paternalism and is merely 'old wine in new casks'. Although he showed successfully that the similarities between regular paternalism and libertarian paternalism are more important than what Sunstein and Thaler (2003) sometimes suggest, my argument here shows that there is nevertheless an important qualitative shift when moving from regular paternalism to libertarian paternalism: libertarian paternalism does not interfere with people's liberty,[14] but with their autonomy, and render possible some kinds of justification that would not be available to regular paternalism. For example, a full ban on smoking would be a (non-libertarian) paternalistic measure and would infringe on people's liberty to smoke; it would be difficult, in order to determine the ethicality of such a law, to balance the loss of liberty versus the gain in health and autonomy for the addicted smoker. On the contrary, health warning messages on tobacco products interfere with smokers' autonomy in order to foster (amongst other goods) this very same autonomy. Here, the loss and gains are easier to put in balance because they concern the same good: autonomy. Therefore, the case of health warnings on tobacco products illustrates that classically paternalistic and libertarian paternalistic measures differ in an important aspect. This shows how the study of a specific applied problem can shed light on a quite general and fundamental debate in ethics and political philosophy.

Acknowledgements

The author would like to thank two anonymous reviewers for their very helpful comments, as well as attendance at the philosophy seminar of the Division of Philosophy at KTH for their useful suggestions—including Till Grüne-Yanoff, William Bülow, Payam Moula, Erik Dahl, John Cantwell and Karim Jebari—and interesting inputs from private discussions.

The ideas expressed in this article are solely the opinions of the author and do not necessarily represent the opinions of the funding agencies.

Funding

This work was supported by the Swedish Institute and the Japanese Society for the Promotion of Science.

Conflict of Interest

None declared.

Notes

1. The dominant view in the literature is that health warnings are an effective tobacco-control strategy;

but see e.g. Peters *et al.* (in press) for an alternative view.

2. Many nudges aim at triggering non-deliberative faculties, like the food arrangement mentioned above. However, as was pointed by an anonymous reviewer, some of them aim at blocking non-deliberative faculties, like mandatory cool-off periods before making an important expanse, which may enhance opportunities for deliberation. It is to my knowledge an open question whether nudges are coextensive with interventions that interact with non-deliberative faculties in some way (by either triggering them or blocking them).
3. At least, according to some classical accounts of liberty, but see e.g. Grüne-Yanoff (2012) who shows that they are not liberty-preserving according to Isaiah Berlin's (2002) account of liberty.
4. The account of what it means to 'address' non-deliberative faculties will presumably depend on the specific account of non-deliberative faculties that is chosen. A tentative definition would be that an intervention addresses non-deliberative faculties when the processing of the information conveyed by this intervention is made (at least in part) by non-deliberative faculties (and when this was intended by the person who designed the intervention).
5. The underlying cognitive mechanism has sometimes been called the 'affect heuristic' (Slovic, 2000).
6. Even if the bias in the lottery case is more significant, as it is unfortunately much more likely to get Brueger's disease when smoking than winning the first prize when playing lottery.
7. The lottery advertisement is actually less misleading than this tobacco warning message, as it is designed in a humorous way: virtually everyone reading the advertisement can realize that confusing it with the statement '100% of people who tried their luck have won' would be blatantly false.
8. Most smokers are addicted: it is estimated that 90% of smokers would like to stop smoking without success; and over 90% of attempts to quit smoking fail in the first year (Carlson and Luhrs, 1997). Moreover, most smokers are unwilling addicts—if they could smoke without being addicted, they would choose that option. Therefore, the phrase 'typical smoker' will refer to an unwillingly addicted smoker. Two marginal cases, the non-addicted smoker and the willing addict, will be considered in a later section.
9. Even if the warning message would reach the inside of the cigarette pack, it would be easy to transfer the cigarettes into a warnings-free case.
10. 'Substantially' only, because the smoker will still be exposed to the warnings every time she buys cigarettes or removes a pack from the case, as pointed by an anonymous reviewer.
11. It is an open question whether this independence component has any value in itself (that is, a value associated with its procedural dimension), beyond the value of leading to authentic choices when it does so (that is, a value associated with its non-procedural dimension). Here, I just make the reasonable assumption that even if this component has any value in itself, this value is of lesser extent than the value of leading to authentic choices. This may differ from other standard hierarchical accounts of autonomy that do not balance the values of the two components one with another.
12. A full definition of authenticity would exceed the scope of this article, but one can adopt here the tentative following definition: an action is authentic on the condition that it would have been chosen by the agent if he would have been in the same state of knowledge as he is, with the same faculties of judgment and decision, but free of any external influence. In our present case, the addicted smoker who is influenced not to smoke (or to smoke less) by the persuasive role of tobacco health warnings is taking an authentic action (because, if he would have been free of any influence from tobacco addiction and persuasive health warnings, he would not have smoked or would have smoked less).
13. One could wonder how authentic would be the choice not to smoke for an addicted smoker who would follow his second-order desire to desire not to smoke. As a matter of fact, one could wonder whether this second-order desire may have originated in the persuasive role of health warnings; in that case, persuasive health warnings would be a kind of propaganda that changes people's second-order desires. However, the smoker would presumably hold this second-order desire even without any persuasive health warnings: indeed, before the introduction of such contemporary, persuasive health warnings, at a time when health warnings were mainly informative, a majority of smokers already wanted to stop or to reduce smoking (cf. e.g. Goodin, 1989). Therefore, if an addicted smoker would chose not to smoke, this choice would be authentic, given the account of authenticity proposed here above.

14. It does not decrease liberty in the classical sense of keeping all alternatives open; as mentioned earlier, Grüne-Yanoff (2012) argues that it does reduce liberty according to Berlin's account.

References

Berlin, I. (2002). *Liberty*. Oxford: Oxford University Press.

Bovens, L. (2009). The Ethics of Nudge. In Grüne-Yanoff, T. and Hansson, S. O. (eds), *Preference Change: Approaches from Philosophy, Economics and Psychology (Theory and Decision Library A)*. Berlin and New York: Springer, pp. 207–220.

Carlson, M. and Luhrs, C. (1997). The Ethics of Tobacco Marketing. In Burkhart, L., Friedberg, J., Martin, T. and Sharma, K. (eds), *Confronting Information Ethics in the New Millennium*. Boulder, Colorado: Ethica Publishing, pp. 72–78.

Chapman, S. (1996). The Ethics of Tobacco Advertising and Advertising Bans. *British Medical Bulletin*, **52**, 121–131.

Chapman, S. and Lieberman, J. (2005). Ensuring Smokers Are Adequately Informed: Reflections on Consumer Rights, Manufacturer Responsibilities, and Policy Implications. *Tobacco Control*, **14(Suppl 2)**, ii8–ii13.

Christman, J. (2011). Autonomy in Moral and Political Philosophy. In Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, available from: http://plato.stanford.edu/archives/spr2011/entries/autonomy-moral/.

Dworkin, G. (2010). Paternalism. In Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*, available from http://plato.stanford.edu/archives/sum2010/entries/paternalism/.

Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *Journal of Philosophy*, **68**, 5–20.

Goodin, R. E. (1989). The Ethics of Smoking. *Ethics*, **99**, 575–624.

Gregory, R. and Mendelsohn, R. (1993). Perceived Risk, Dread, and Benefits. *Risk Analysis*, **13**, 259–264.

Grüne-Yanoff, T. (2012). Old Wine in New Casks: Libertarian Paternalism Still Violates Liberal Principles. *Social Choice and Welfare*, **38**, 635–645.

Hammond, D. (2011). Health Warning Messages on Tobacco Products: A Review. *Tobacco Control*, **20**, 327–337.

Hansen, W. B. and Malotte, C. K. (1986). Perceived Personal Immunity: The Development of Beliefs About Susceptibility to the Consequences of Smoking. *Preventive Medicine*, **15**, 363–372.

Hausman, D. M. and Welch, B. (2010). Debate: To Nudge or Not to Nudge. *Journal of Political Philosophy*, **18**, 123–136.

Kruglanski, A. W. and Gigerenzer, G. (2011). Intuitive and Deliberate Judgments are Based on Common Principles. *Psychological Review*, **118**, 97–109.

Marsh, A. and Matheson, J. (1983). *Smoking Attitudes and Behaviour*. London: Her Majesty's Stationery Office.

McCoy, S. B., Gibbons, F. X., Reis, T. J., Gerrard, M., Luus, C. A. and Sufka, A. V. (1992). Perceptions of Smoking Risk as a Function of Smoking Status. *Journal of Behavioral Medicine*, **15**, 469–488.

McKenna, F. P., Warburton, D. M. and Winwood, M. (1993). Exploring the Limits of Optimism: The Case of Smokers' Decision Making. *British Journal of Psychology*, **84**, 389–394.

Mill, J. S. (1859). *On Liberty*. London: Penguin Books.

Pears, D. (1984). *Motivated Irrationality*. Oxford: Clarendon Press.

Peretti-Watel, P., Halfen, S. and Grémy, I. (2007). The 'Moral Career' of Cigarette Smokers: A French Survey. *Health, Risk & Society*, **9**, 259–273.

Peters, G.-J. Y., Ruiter, R. A. C. and Kok, G. (2012). Threatening Communication: A Critical Re-analysis and a Revised Meta-analytic Test of Fear Appeal Theory. *Health Psychology Review*, doi:10.1080/17437199.2012.703527.

Rizzo, M. and Whitman, D. (2008). Little Brother is Watching You: New Paternalism on the Slippery Slopes. *New York University Law and Economics Working Papers*, Paper 126, available from: http://lsr.nellco.org/nyu_lewp/126 [accessed 26 February 2013].

Schoenbaum, M. (1997). Do Smokers Understand the Mortality Effects of Smoking: Evidence from the Health and Retirement Survey. *American Journal of Public Health*, **87**, 755–759.

Schooler, L. J. and Hertwig, R. (2005). How Forgetting Aids Heuristic Inference. *Psychological Review*, **112**, 610–628.

Slovic, P. (2000). *The Perception of Risk*. London: Earthscan Publications.

Sperber, D. and Wilson, D. (1995). *Relevance: Communication and Cognition*. 2nd edn. Oxford: Blackwell.

Sunstein, C. R. and Thaler, R. H. (2003). Libertarian Paternalism is Not an Oxymoron. *University of Chicago Law Review*, **70**, 1159–1202.

Thaler, R. H. and Sunstein, C. R. (2008). *Nudge. Improving Decisions About Health, Wealth and Happiness*. New Haven: Yale University Press.

Tversky, A. and Kahneman, D. (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, **5**, 207–232.

Villejoubert, G. and Mandel, D. R. (2002). The Inverse Fallacy: An Account of Deviations from Bayes's Theorem and the Additivity Principle. *Memory & Cognition*, **30**, 171–178.

Viscusi, W. K. (1990). Do Smokers Underestimate Risks? *The Journal of Political Economy*, **98**, 1253–1269.

Voigt, K. (forthcoming). 'If You Smoke, You Stink.' Denormalisation Strategies for the Improvement of Health-Related Behaviours: The Case of Tobacco. In Strech, D., Hirschberg, I. and Marckmann, G. (eds), *Ethics in Public Health and Health Policy*. Amsterdam: Springer.

Weinstein, N. D., Marcus, S. E. and Moser, R. P. (2005). Smokers' Unrealistic Optimism About Their Risk'. *Tobacco Control*, **14**, 55–59.