Automatic Actions – Agency, Intentionality, and Responsibility

Christoph Lumer
University of Siena, Italy

Published in: Philosophical Psychology. 29 pp. Preprint online since 13.3.17, URL: <http://www.tandfonline.com/doi/full/10.1080/09515089.2017.1291928>, 13.3.2017. DOI: 10.1080/09515089.2017.1291928. Permanent link: http://dx.doi.org/10.1080/09515089.2017.1291928.
(Orthography and punctuation of the printed version, with respect to the present version, have been aligned to the publisher's style.)

Abstract: The article discusses a challenge to the traditional intentional-causalist conceptions of action and intentionality as well as to our everyday and legal conceptions of responsibility, namely the psychological discovery that the greatest part of our alleged actions are performed automatically, i.e. unconsciously and without a proximal intention causing and sustaining them. The main part of the article scrutinises several mechanisms of automatic behaviour, how they work and whether the resulting behaviour is an action: actions caused by distal implementation intentions, four types of habit and habitualisation, mimicry and semantically induced automatic behaviour (which, however later is disregarded because of its unclarity). According to the intentional-causalist criterion, the automatic behaviours resulting from all but one of these mechanisms turn out to be actions and intentional; and even the behaviour resulting from the remaining mechanism (naturally acquired habits) is something we can be responsible for. Hence, the challenge, seen from close up, does not really call the traditional conception of action and intentionality into question.

Keywords: automatic behaviour, habits, mimicry, auto-motives, intentional causalism, intentionality, agency, responsibility

1. The Topic: Automatic Behaviour, Agency and Responsibility

The most widely accepted classical conception of action is *intentional causalism*: An *action* (in the strict sense) is a behaviour caused (in a non-deviant way [Lumer 2008]) by a respective intention; and to be really ours the *intention* has to be volitive, i.e. a (in principle) rationalisable integration of our conscious attitudes towards the action, and hence conscious or preconscious. The philosophical reason for intentional causalism, apart from its (arguable) empirical reality [cf. Lumer 2007], is that it explains the value of actions, namely that actions conceived in this way give practical power to our conscious ego: make it control our behaviour and thereby change pieces of the world for realising our desires and implementing the decisions of our reason. Conscious intentions are pivotal in this conception. On the one hand, they represent the upshot of the subject's reflection, the balancing of his desires, the integration of his various attitudes towards the options and express the concerns of the person and thereby also the personality itself. On the other, intentions have executive power to realise the intended behaviour and piece of the world by causing the intended behaviour. [Lumer 2013.] The intentional-causalist conception has not only been dominant in the history of philosophy – from Aristotle via Hume, Kant to e.g. Davidson – it is also prevalent in

everyday thinking of actions, and it is the basis of criminal, civil as well as moral responsibility in its usual meaning and hence also of the criminal justice system (cf. e.g. Kenny 1978; Vargas 2013).

Now, recent findings in cognitive and neurosciences have challenged this picture in various ways. One of these challenges is that most of our actions are automatic, hence not at all caused by a proximal intention. The article discusses this challenge: What kinds of automatic behaviour are there? How do they function? Is the resulting behaviour still intentional and, therefore, still an action in the strict sense? Are we or can we be responsible for our automatic behaviour?

In the following, first, the challenge is explained more in detail (sect. 2). Next, considering the topic of the present article, a working definition of 'automatic behaviour' is developed, which delineates the object under study (sect. 3). The article's main part then distinguishes various types of automatic behaviour, analyses the underlying mechanisms [1] and, on this basis, assesses the potential intentionality and the agential character of the resulting behaviour as well as our potential responsibility for it (sects. 4-7). Then some alternative philosophical explanations of automatic behaviour are criticised (sect. 8). After recapping that much of automatic behaviour is still intentional and that we are also responsible for a large part of the non-agential automatic behaviour, the questions of intentionality and responsibility are examined in more detail, in particular by differentiating various types of responsibility and by answering objections. If we are responsible for so much of our automatic behaviour this is a chance but also a burden (sects. 9-10).

2. The Challenge of Automatic Behaviour for Agency and Responsibility

Most of our actions are *automatic* in the sense of being triggered by a consciously or unconsciously perceived signal that sets in motion an automatism of executing certain actions, where this automatism works smoothly and effortlessly and does not require conscious attention [cf. e.g. Bargh & Chartrand 1999, 463; Neumann 1984, 282]. Someone scratches his head or twists his beard; a driver shifts gears at a certain pitch or she drives while listening to the news; a daughter is telling long stories to her father while taking a walk; a trained typist is talking to her colleague while copying a text [Kihlstrom 1987, 1447]; a writer sets down the single words of a sentence while still pondering whether the formulation is good. Some psychologists say that only 5% of our actions are consciously controlled, the rest is automatic [Baumeister et al. 1998, 1252; Baumeister & Sommer 1997]. This holds for that reason alone that that the majority of adults' actions are molecular actions, which are composed of smaller molecules and finally atomic actions, where the atoms are triggered and executed automatically.

But are seemingly automatic actions really without intention or at least not triggered by a proximal intention? At first glance this seems hard to believe. So let us consider some empirical findings. It is common knowledge and the basis of the respective practice in learning, e.g. to drive a car or to play a musical instrument, that we can unburden our attention by repeating and training

1 "Mechanism" here is meant in the sense of: "the fundamental fixed causal processes responsible for some (repeating) phenomenon" [cf. Babcock Gove 1993, mechanism 4.a].

smaller chunks of behaviour thus *automatizing* it so that we are later able to use our attention for other tasks [Bargh & Chartrand 1999, 468]. And it is part of common knowledge that if we later focus our attention again on the details, the overarching action often is severely disturbed – it can e.g. be dangerous to run down a staircase and in the middle to start to mind one's single steps. Several sorts of *action slips* (Norman 1981) are another everyday evidence for the lack of proximal conscious intention in many actions: Ambiguous stimuli of automatic actions may trigger the false action precisely because the triggering is not consciously controlled – an urban bus driver driving his private car along his bus route may swing out to stop at the bus stop [Heckhausen 1987, 157], somebody going into her bedroom to dress for dinner may end up in bed for sleeping [James 1890, I, 115]. *Bad habits* are further evidence of common knowledge for the lack of conscious intentions in automatic actions as they may persist against our prior intention to stop them. Psychological research has revealed that bad habits, unlike responses to temptations, are controlled most effectively through spontaneous use of vigilant monitoring (thinking "don't do it," watching carefully for slip-ups) [Quinn et al. 2010, 499]. This means, that precisely because there is no conscious attention to the topic the bad habit can do its work. The evidences for the lack of conscious (proximal) intentions cited so far are still indirect and one may question them. Therefore, some empirical research tries to show that in several cases there cannot be a respective conscious proximal intention for automatic action. One way of doing this is to ensure that not even the trigger gets conscious before the action, so that the subject still less can form a proximal conscious intention as a reaction to it. The simplest cases are fast reactions like *starting to sprint:* Professional sprinters start to run before they can have consciously heard the starting shot: the lapse of time for building the conscious perception of a signal, as established in other experiments (300-500 ms), is longer than the sprinter's objectively measured reaction time (50-100 ms) [Libet 1985, 559]. Probably there was a distal conscious intention to 'run immediately after the starting shot' but not a proximal intention, formed after the shot. Another way, used in the lab, to show the lack of a proximal intention is to operate with an unperceivable trigger – as in experiments demonstrating the *Fehrer-Raab effect:* The trigger (e.g. a black disc on a screen) is shown only very briefly and masked by another, longer displayed signal (e.g. a black ring) so that the trigger will never be conscious – afterwards persons say that they have seen the masking signal (ring) but not the trigger (disc). In the experiment persons were told to press a button as quickly as possible after seeing "the" signal. Now the time between the beginning of the trigger signal and the beginning of the masking stimulus varied but the reaction time between the trigger signal and the pressing remained constant (mean 160-165 ms) and roughly equal to the reaction time in a control experiment with one stimulus only. [Neumann & Prinz 1987, 201 f.] The most obvious explanation of the Fehrer-Raab effect is: (After forming a distal general intention 'to press the button immediately after the signal' at the beginning of the experiment when the experimenter explains what the subjects should do) the subjects reacted directly to the consciously unperceivable trigger signal; and this excludes that in between, after the signal, they first formed a (triggering) conscious proximal intention. A final evidence for the lack of proximal intentions in habitualized automatic actions is *neurophysiological*: Stimulus-response routines are stored in the basal ganglia, in particular in the

putamen, whereas normal goal-directed actions are controlled in other brain regions, often including the prefrontal cortex [Neal et al. 2011, 1429 – with several references].

The main functional reason for automaticity is that automatic actions (and automatic processes in general) do not consume the very scarce resource of conscious attention so that during the execution of the automatism these higher psychic capacities can be used for other tasks [Posner & Snyder 1975; Bargh & Chartrand 1999, 464]. As a consequence, automatic reactions are mentally effortless. Furthermore, they are much faster than deliberate conscious reactions. Finally, automatic behaviour can be executed in parallel without much interference [Neumann 1984, 260-264]. Generally mentioned disadvantages of automatisms, however, are, first, that they fail in complicated processes, second, that they cannot operate in new situations (in new situations execution problems often arise, which then attract attention so that the problem can be resolved via conscious thinking), third, that they operate rigidly and mechanically even when deviating from the rigid rule would be better, e.g. when the automatic behaviour is no longer rewarding, and, fourth, that they are susceptible to action slips [Neal et al. 2011, 1429; Graybiel 2008, 363; Reason 1979].

So it is part of the very nature of automatic actions that they are triggered and proceed unconsciously; they are driven by stimuli not by present intentions [Neumann 1984, 258]. This could be an enormous challenge to the traditional picture and treatment of actions:[2] Without intention there is no reason for which the action has been done, no intentionality, no freedom, no responsibility; and in the strict sense there is not even an action; and this would hold for the vast majority of what appear to be actions.[3] Furthermore, without the interposition of critical and constructive conscious processes not only is no creative designing of new options possible, but also the search for all kinds of relevant consequences – which requires consciousness as the global working place – and the rational assessment of these consequences are omitted [Lumer 2014, 86-87; 96-100); we are behaving on autopilot and not as agents. Finally, automatisms, and bad habits in particular, can even work against our deliberate intentions; we have to suppress them with proper conscious efforts [Quinn et al. 2010; Neal et al. 2011, 1436]. Of course, all these features would be severe curtailments of rational agency; and some of them may exclude our responsibility.

To make the challenge by automatic actions for intentionality and responsibility more concrete let us modify the above examples, which show the absence of (proximal) intentions from automatic actions. A *fare dodger* habitually uses public transport without paying and circumvents

---

2 These days, several psychologists and neuroscientists (e.g. Benjamin Libet [1985], Daniel Wegner [2002, 2; 65-69; 96; 146; 342], Gerhard Roth [2001, 441-442], Christof Koch, Francis Crick [Koch & Crick 2001]) deny a decisive role of intentions for our actions in general. I have replied to such challenges elsewhere [Lumer 2014a; 2014b]. But there are also intention sceptics who rely specifically on the discovery of actions' automaticity. Psychologist John A. Bargh, e.g., takes the discovery of the pervasiveness of automatic actions to be a confutation of traditional conceptions of action (as well as of free will) [Bargh 2005; 2008]. (His reason for this appraisal is that if automatic actions are caused by intentions at all these intentions, according to his theory, are unconscious and automatic as well. Bargh's theory is discussed below (sect. 7).)

3 In Western penal law, guilt e.g. presupposes intent or negligence; furthermore, it presupposes *mens rea*, i.e. the knowledge of doing something forbidden, which implies conscious knowledge of what one is doing. The reason for these conditions is that threat of punishment can be deterrent only for deeds fulfilling them.

possible mechanical controls routinely without conscious thinking. A *mafioso* has placed a bomb under the street to blow up an enemy in his car; the mafioso is waiting with his remote control in a place where he has only a view through a narrow slot on the place with the bomb; an accomplice watching the street from another place tells him that the enemy is approaching; the mafioso now waits for the next car which can be seen through the slot and presses the trigger as fast as possible when he acknowledges a car. The fare dodger and the mafioso could both say that they did not act on a (proximal) intention but automatically; hence they did not fare-dodge or kill intentionally. If that were true the fare dodger under many legislations could not be punished at all; and the mafioso could not be punished for murder or manslaughter. A different kind of example is this: Some *reckless driver* habitually drives too fast or in other ways recklessly (e.g. ignores the right of way) and one day causes an accident. The usual accusation in such cases is *not* that the driver caused the accident intentionally, but that she caused it at least negligently; the accusation often is not even that the driver drove too fast or ignored the right of way intentionally but that she *omitted* to curb the speed or to check for others who might have the right of way and then give way where required. Because the charge is that of an omission the challenge by the automaticity of actions in this case is *not* that the driver might not be responsible because she did something without intending it. But there is a related problem. The driver could say: 'Driving a car (like many other activities) is mainly automatic; i.e. we cannot intentionally control these activities, at least not for the most part; trying to intentionally control them would mean to give them up altogether. Hence it is impossible to do the omitted action intentionally; I could not act in the required way, and, therefore, I am not responsible either for the accident or for going too fast / driving recklessly. If others do not cause accidents etc. they are simply lucky e.g. because they have better automatisms.' A final example concerns a different domain of responsibility. If somebody is *obese* because she automatically ingests all sweet things in reach this is not a problem for legal (penal or civil) responsibility because this is not a behaviour regimented by law. However, this automaticity may generate a problem for personal, prudential responsibility: If the obese person does not eat intentionally she is not prudentially responsible for this part of her life.

The most general hypothesis in what follows is that in automatic behaviour there is much more intention, controllability and responsibility as first it appears to be, with the consequence that most of this behaviour is intentional. Some psychologists also see such a connection between automatic behaviour and intention. Wood & Neal e.g. write: "Nonetheless, habits interface with goals. Constraining this interface, habit associations accrue slowly and do not shift appreciably with current goal states or infrequent counterhabitual responses. Given these constraints, goals can (a) direct habits by motivating repetition that leads to habit formation and by promoting exposure to cues that trigger habits, (b) be inferred from habits, and (c) interact with habits in ways that preserve the learned habit associations" [2007, 843]. Neumann states that automatic processes are not independent of a person's intentions and attention [1984, 256]. So there is even explicit support from empirical scientists for the general hypothesis elaborated below. However, this elaboration itself, i.e. the detailed analysis of the exact role of intentions in automatic behaviour and the explanation of why and how far automatic behaviour is intentional, is new and the constructive

contribution of this paper.

3. Circumscribing the Object under Scrutiny – a Definition of 'Automatic Behaviour'

There is some discussion in psychology about how 'automatic behaviour' should be defined. This article does not aim at presenting a substantial contribution to this discussion and will use only a working definition, which on the one hand takes up the most important qualities proposed by psychologists and, on the other, is suitable for the present discussion in not predetermining the question of the intentionality and agential character of automatic behaviour. This working definition is:

> *Automatic behaviour* is a (1) behaviour of a kind which in principle is (directly) voluntarily controllable, (2) which is unconsciously triggered, (3) brought about in a (more or less) reliable way, (4) not initiated by a (proximal) intention, and (5) not consciously controlled during its execution.

The condition of basic voluntary controllability (1) – which does not imply *present* voluntary controllability – shall exclude philosophically unproblematic cases (like reflexes), where, for physiological reasons alone, no confusion with actions is possible. Unconscious triggering (2) is one of the core conditions of automaticity; it permits the trigger itself to be conscious – we may e.g. consciously and attentively perceive a certain signal –; but the subsequent process of *triggering* must not be conscious, it must be automatic – otherwise "automatic" behaviour would be too close to paradigmatic cases of deliberate conscious action. Without reliability (3) the behaviour might be accidental and hence not automatic. Not being initiated by a proximal (conscious or unconscious) intention (4) again excludes paradigms of intentional action [4] and is a requirement of automaticity: The very idea and function of intention (in a narrow sense [cf. Lumer 2013]) formation with its binding to the subject's desires, its openness to creativity, critical scrutiny and search for relevant consequences as well as alternatives preclude its automatisation. Hence permitting automatic behaviour to be initiated by a proximal intention, paradoxically would allow "automatic" behaviour to be initiated by a non-automatic process. (This does not exclude that unconscious executive states that are not intentions in a narrow sense are formed automatically and produce automatic behaviour (cf. below, sect. 7).) Finally, lack of conscious control during the execution (5) is the other core condition of automaticity. – To be sure, every consciously controlled action includes much automatic processing, after all the conscious ego does not think of giving efferent signals to

4 'Intentional action' has not the same meaning as 'action'. Though every action must be caused (in the right way) by an intention, sometimes actions fail so gravely that no underlying intention is realized. In such a case the action is not intentional under any description – though it may still be an action because it has been caused by an intention via an action generating mechanism. However, actions which are not intentional under any description are rare. So, when I write "intentional action" here (without specifying the content of that intention) I only want to stress the most important feature of normal actions, viz. that they realize the causing intention (or at least some part of it) in the right way.

motoneurons, to inhibit counter-acting muscles etc. But, according to the terminology chosen here, these are automatic *processes* not forms of automatic *behaviour*. The question about automatic behaviour is who has the leadership, and not whether that leader consciously controls the details of the execution. This means: The very details of our actions are not consciously controlled – as in a big company the CEO does not know the single operations of the workers –; this is a truism and not an issue in the debate on automaticity. The question is whether at the top of the executive hierarchy there is a conscious intention or an unconscious automatic process [cf. Wilson 2002, ch. 3: Who's in charge?].

When psychologists speak of automaticity they usually speak of 'automatic *processes*', which is a broader concept than 'automatic *behaviour*' as it has been defined here: First, the *genus proximum*, 'process', is much broader than 'behaviour', which, according to condition 1 of the definition (voluntary controllability), includes only such processes that at least in principle could be actions. Second, condition 4, "not initiated by a (proximal) intention", is motivated by the aim of the present philosophical inquiry, namely to discuss behaviour which might be problematic for intentional causalism because it is not caused by a "visible" intention though it seems to be an action; so this condition, too, makes the concept 'automatic behaviour' narrower than the psychological concept 'automatic process'. The remaining three conditions of the definition, instead, are broadly accepted by psychologists as well. As a consequence of these determinations, the above definition excludes several phenomena that some psychologists include in their discussions of automaticity, but which, from the viewpoint of the present philosophical discussion belong to quite different categories of, perhaps again, problematic, cases. The most important of these excluded phenomena are the following. *1. Conditioned reflexes* – like closing one's eye automatically after an acoustic signal, if one has previously, several times and regularly been exposed to air jets after this acoustic signal – are not automatic behaviour because the underlying reflex (closing the eye when feeling the air jet) is not a consciously controlled action in the first place. (Implicitly) learned extensions of the conditions under which these reflexes are triggered do not make them actions. *2. Primed actions* are influenced by automatic processes which intervene on the process of intention formation or on the mode of the action's execution, but they are not forms of automatic behaviour because they are initiated by conscious proximal intentions (cf. condition (3)) and then are, of course, actions. Not even the intention formation underlying a primed action is an automatic process since there is only *some* influence on the decision and not a real automatism leading to a stereotypic intention. *3. Unconsciously deliberated actions* are unconsciously caused by an unconscious intention, which in turn relies on a primitive unconscious deliberation. Since the unconscious deliberation leads to an unconscious *intention* the resulting behaviour, according to the above definition, is not automatic.

Having excluded these three types of behaviour from the present discussion does not mean that they are not problematic for intentional causalism. The present point, instead, is only that they are not problematic for action philosophy with respect to their possible *automaticity*. If they are problematic – and in particular primed actions as well as unconsciously deliberated actions prima facie seem to be so – they are problematic for other reasons and have to be and are discussed in

other studies (cf. Lumer, under review). On the other hand, the definition of 'automatic behaviour' does include some processes which are not actions and commonly are also considered to be clear cases of non-actions. Therefore, the definition, correctly, does not imply that automatic behaviour always is also an automatic *action*; thus the definition leaves our research question, i.e. whether certain kinds of automatic behaviour are actions, open. An example for automatic behaviour (in the defined sense) that is not an action is *routine respiration*. Routine respiration fulfils the four standard conditions of automaticity (it is 2. unconsciously triggered, 3. brought about reliably, 4. not initiated by a (proximal) intention, and 5. not consciously controlled during its execution) but also the first condition (voluntary controllability), since we can voluntarily modulate the rate and volume of each breath we take. Hence, routine respiration is an automatic *behaviour*. But is it also an automatic *action*? If this were the case respiration could be the model for explaining other types of automatic actions. However, the fact that respiration is voluntarily controll*able* does not imply that it is voluntarily *controlled* in each case; and *routine* respiration precisely is not voluntarily controlled. Furthermore, routine respiration does not originate from some respective intention, not even a distal intention. Therefore, it is not intentional, not an action, and we are not responsible for it. Thus, automatic respiration cannot be the model for explaining automatic actions. The situation is different with consciously controlled respiration, i.e. when we consciously focus on our breathing and intentionally modulate it: Now it is consciously triggered by an intention, such that conditions 2 and 4 of the above definition are no longer fulfilled; therefore, consciously controlled respiration is no automatic behaviour. Instead it is an (intentional) action because it is caused (in the right way) by a respective intention; and we are responsible for it. So, both kinds of behaviour are categorized by our definitions as they should be: Routine respiration is taken as a classical paradigm case of mere (automatic) behaviour (i.e. not an action), whereas consciously controlled breathing is considered an action.[5] Therefore, if the common assessment of the kind of automatic behaviour under scrutiny in this article – like automatic walking, driving etc. –, i.e. that they are actions, were correct, this behaviour must be different from respiration in some important respect. And the idea is that this automatic behaviour, the first impression notwithstanding and in contrast to routine respiration, is caused by a somewhat hidden respective intention.

A similar but more complicated case is conscious mental processes. It is more complicated because we have do differentiate: 1. Some mental processes are directly intentionally controllable, though often they are not intentionally controlled: e.g. directing and focusing attention, certain mental representations (in particular visual and auditory), judging etc. 2. Others are not directly intentionally controllable but indirectly with the help of the first group, i.e. directly controllable mental processes, and via mechanisms which remain in the mental sphere: e.g. emotions with the help of suitable representations; creative inspirations with the help of mentally representing the question or (partial) ideas for their solution; memorization with the help of mental repetition, concentration and the construction of mnemonic aids etc. 3. And a final group of mental processes

5 By the way, the same holds for suppressible reflexes like the nocifensive reflex to withdraw our hands when we touch something hot. Usually this is a kind of automatic behaviour and not an action. But the reflex can be consciously and intentionally suppressed; then the standstill is an action and not an automatic behaviour.

are intentionally controllable only via intervention on the environment: e.g. the content of our sensory perceptions. Groups 2 and 3 do not fulfil the controllability condition 1 of the definition of 'automatic behaviour' and, therefore, are not interesting here. Group 1 instead is relevant in the present context but has to be broken down further: 1.1. Many of the processes which are of the directly intentionally controllable type in fact are not controlled: our attention is attracted involuntarily; ideas, images, phrases pop into our heads etc. Such a process might be automatic; and because of its controllability in principle it might be automatic behaviour in the defined sense. However, it is not an action because it lacks the intention. This is analogous to routine respiration. 1.2. If such a process is triggered by an intention instead it is an action. This is analogous to intentional respiration. 1.3. Finally, it is possible to always perform such processes intentionally in certain situations, which may lead to habitualization and automatization of these processes. Such habitualizations are even part of our education, when we are taught to follow certain rules of mental behaviour: 'Before crossing the road, first look left, then look to the right!' 'If you want to read a text start at the left side of the top line …!' 'If you translate a Latin sentence first look for the verb!' 'If you are taking a difficult decision first realize what the relevant options are …' 'If you are undecided between two options and do not make progress with the decision for some time, and if the temporal costs of the decision are no longer proportioned, then take a spontaneous decision in that instance (or postpone the decision)!' What at the beginning is a series of intentional mental actions, at the end of the habitualization has become a habit without proximal intentions. This implies that the resulting habitual behaviour is automatic behaviour in the defined sense. Whether this or other habitual behaviour is an action is an open question, which will be addressed below, in section 5.[6]

4. Automatic Behaviour, Type 1: Actions Caused but not Triggered by Distal Implementation Intentions

Now, the middle part of this article will discuss the main forms of automatic behaviour: 1. actions caused but not triggered by distal implementation intentions, 2.-5. four forms of habits, 6. mimicry, and 7. semantically induced automatic behaviour.

Gollwitzer has distinguished goal intentions from implementation intentions [Gollwitzer 1999]. Both kinds of intentions are, qua intention, intentions to execute an *action*, i.e. to do something; however, the actions in the two kinds of intentions are described differently. In a *goal intention* they are described via the goal to be reached, i.e., logically conceived, as 'to execute some behaviour *x* which causes the achievement of goal *g*' (where "*x*" is a variable and "*g*" is a singular

6 This analysis and differentiation of mental behaviour also entails an answer to the question of the controllability of mental actions, to which some answer in the positive [e.g. Wu 2013], others in the negative [e.g. Strawson 2003; Vierkant 2013], and some say in part yes, in part no [e.g. Hieronymi 2009; Mele 2009]. Furthermore, the distinction between spontaneous judgements or decisions (group 1.1) on the one hand and intended judgements or decisions (group 1.2) on the other resolves the problem of a possible infinite regress while leaving the possibility of intentionally deciding to decide intact.

term), whereas in an *implementation intention* the action's parameters are specified in a way that they are understandable for the executive system; what is understandable in this way varies with the degree of practice. As a consequence, implementation intentions can be directly executed, whereas in order to realise the goal intention an appropriate implementation intention has to be formed first. Goal intentions often are distal intentions to be executed in the not immediate future, implementation intentions instead paradigmatically are proximal intentions to be realised right now. However, there are also distal implementation intentions, i.e. intentions whose time index in the action description is not 'right now' but '(immediately after) when event *e* occurs' with *e* being some event in the more distant future: 'I do action *A* (immediately after) when *e* occurs'. And the interesting critical feature here is that distal implementation intentions can be executed automatically, i.e. in particular without first forming a proximal implementation intention; the mere observation of *e*, even the unconscious perception of *e*, is sufficient to trigger the *A*-ing. Brandstätter et al. [2001, experiment 3] have shown this with laboratory experiments, which were specifically designed to occupy the participants' attention with other absorbing tasks and in which the specific signals to which the subjects, according to their interesting distal implementation intention, should selectively react appeared outside of awareness (in the parafoveal area of the visual field, for which it holds that the stimuli presented there "are processed only minimally and outside of awareness"). The reaction time after these signals remained the same whether the absorbing task was difficult or rather easy – which the authors take to be another proof of the automaticity of the distal intention's execution. An everyday example is to plan to post a letter on one's way to work, e.g. holding it ready in the right hand; when one passes the mailbox one inserts the letter automatically. Some further examples from the literature of unconscious and automatic triggering the execution of distal intentions have already been described in section 2: starting to sprint, Fehrer-Raab effect.

The mechanism behind distal implementation intentions' automatically causing the intended behaviour must be a bit more complex than that of proximal implementation intentions: The distal intention – like a proximal intention – is still the behaviour's structuring cause, thus determining how we react on the perception of which signal *e*. However, the intention no longer triggers the intended behaviour but causes a programming of the execution mechanism to trigger the behaviour after the perception of event *e* (signal) mentioned in the action description's time index – like setting an alarm-clock or the timer of some instrument or placing a conditional selling order for shares when they fall below a certain threshold. Then this programme is executed more or less reliably by the execution mechanism. Many animals are able to learn stimulus-response ties by repetition. In the case of distal implementation intentions we seem to be able to simply *set* such a stimulus-response tie by the distal implementation intention.

According to intentional causalism, actions caused but not triggered by distal implementation intentions are, of course, actions in the full sense because they are caused in a non-deviant way by a corresponding (structuring though not triggering) intention; the definition of 'action' at the beginning of this paper does not say anything about the time lapse between the intention and its execution, thus also permitting later executions. Therefore, normally such actions

are intentional and, usually, we are responsible for them. This shows that the organisation of our actions can be cleverer and more efficient than we usually assume: With the help of distal implementation intentions we can program ourselves to act automatically in the future without having to spend attentional resources in that moment.

5. Automatic Behaviour, Types 2-5: Habits

Habits emerge from the gradual learning of associations between responses and the features of performance contexts that have historically co-varied with them (e.g., physical settings, preceding actions). Once a habit is formed, perception of contexts triggers the associated response without a mediating goal. [Wood & Neil 2007, 843; similar: Shiffrin & Schneider 1977; Bargh & Chartrand 1999, 468.] Initially the actions are caused by conscious intentions, and their performance is consciously controlled as well. However, if the same action is repeated consistently in the same type of situation the conscious control diminishes more and more and finally fades away [details: Neumann 1984, 280-281]; and at a certain point the conscious intention is no longer necessary for causing the now habitualised behaviour [Bargh & Chartrand 1999, 468; Shiffrin & Dumais 1981; Shiffrin & Schneider 1977]. Habitualisation itself, in a certain sense, is an automatic process: The consistent pairing of certain types of situations and actions is sufficient to gradually establish the automatic link between them [Bargh & Chartrand 1999, 469]. As a consequence of its automatisation habitual behaviour is triggered independently of present motives by perceiving the triggering signal if the behaviour can be executed in its habitualised form. [Neal et al. 2011.]

The fact that habitualisation works automatically does not imply that it cannot be influenced. Above all there are various ways in which the original consistent coupling of specific situations and of associated action comes about. Some of them are intentional; and this makes a difference for the potential intentionality and quality of being an action as well as for our responsibility for them. Therefore, in the following subsections the four most important of these ways will be scrutinised separately. In addition to influencing the emergence of habits we can interrupt or delete them if we do not want to have them any longer. Psychologists have found several strategies for achieving this. First, we can monitor the triggering signals and try to inhibit the unwanted response when activated in memory (thinking “don't do it,” watching carefully for slip-ups) [Quinn et al. 2010,]. This strategy, at least for a considerable period of time, requires a constant effort. Second, routines can be interrupted by stimulus control, i.e. the environment is changed in such a way that the precise situation necessary for triggering the habitual response no longer occurs to the subject [Neal et al. 2011, 1436].

After this general characterisation of habits we now have to analyse the four ways in which they are acquired.

5.1. Automatic Behaviour, Type 2: Intentionally Learned Routines and Skills

People, in particular teachers and trainers and, somewhat later, their pupils as well, know – at least roughly – about the automaticity of habitualisation and about the advantages of automatic reactions. Therefore, they often intend to acquire automation and its beneficial features by deliberately repeating and practicing the sequence from the signal to the behaviour until the automatism has developed and stabilised. Musicians, gymnasts, learner drivers do so. [Bargh & Chartrand 1999, 468.] The skills and routines acquired in this way can be independent of a comprehensive action, such as wiping one's nose after sneezing, or inserted into bigger actions and used to fulfil respective subgoals. In the latter case they also fall under the third category of habits (acquiring routines within bigger actions [see subsection 5.3]).

With respect to the possible intentionality, intentionally learned routine behaviour is a simple case. If automaticity acquisition is intended then the agent has a general intention always to perform a certain action type *A* in situations or circumstances of type *C*. This general intention logically extends to every single execution of the automatism. In addition, the instrumental habitualisation or learning intention in the end causes these future executions via the intended practicing. Once the automatism has been acquired, neither the general intention nor the habitualisation intention are triggering causes of the single automatic actions, the latter are triggered by the stimulus perception of the stimulus *C* (which may even be subliminal). However, the general intention, nonetheless, is the structuring cause via the habitualisation intention. Consequently, this kind of automatic behaviour is caused in a controlled way by a respective intention, hence it is an action, intentional, caused by reasons and usually we are responsible for it. But it is a clever form of intentional action, which does not need a singular proximal intention to be executed.

5.2. Automatic Behaviour, Type 3: Habits Acquired by Repetition Under a General Intention

The consistent coupling of the situations *C* and the action *A*, which leads to habitualisation, can also be brought about in a more natural way than just described: We simply have the *general* intention to always *A* in situations *C*; because of this intention we then repeatedly do *A* in situations *C* – perhaps by additionally forming many singular proximal intentions 'now to do *A*' in situations *C* –; and this causes the habitualisation of doing *A* in situations *C*. There is no intention to habitualise the doing of *A* in situations *C*; and we do not undertake any particular actions to establish an automatism. The general intention works and is more powerful than it appeared. Some ordinary examples of this kind of habit may be greeting other persons, thanking for services obtained, or drinking tea at a certain time of day.

Since there is a general intention always to do *A* in situations *C*, which covers all the subsequent executions of *A* in situations *C*, and because this general intention also causes the later habitual executions of *A* via the just described mechanism, these later executions are intended and intentional actions; and usually we are responsible for them. The difference with respect to intentionally learning routines is that the *habitualisation* was not intentional. But this does not infringe on the intentionality of the later automatic *A*-doings. The subject who underwent the unintended habitualisation does not know that much about how his general intention brought about

his later automatic *A*-ings, but he is happy with it, and the way it came about is not surprising for him; in this sense it is not a deviant way of causing the *A*-ings [cf. Lumer 2008]. Hence, these *A*-ings are intentional.

5.3. Automatic Behaviour, Type 4: Routines within Bigger Actions

While the other three types of habits exclude each other, the type of habit to be discussed now, i.e. routines within bigger actions, has to be combined with one of them and it can be combined with each.

As already said, when we often repeat the same action in the same situation the automation takes place even without being intended. A further way in which this can happen is that smaller actions are integrated into bigger and repeatedly executed plans so that single stretches within this bigger plan, which are themselves composed of several smaller actions, can be automated. After the automation the whole plan can be executed in one go and can also be intended *in toto*. One particularity of this setting is that a somewhat more comprehensive intention is part of the identical stimulus *C* which triggers the behaviour sequences; if we do not have this comprehensive intention the elementary actions are not released automatically. As a consequence we have prefabricated action modules, so to speak, at our disposal. These modules can be intentionally activated in isolation or within still more comprehensive plans. This is what happens when we naturally, without any intention to practise, learn to walk, to speak, to close a door, to do up the shoelaces, to enter our house (go to the door, take out the key, …), switch off the light and lock the door when leaving the house.

Is routine behaviour within bigger actions, i.e. the smaller, automatic element like braking or changing gear during one's trip to work, intentional? Is it an action? Even when invited to do so, subjects in many cases cannot list the originally elementary actions involved in the comprehensive action. Nonetheless, we usually assume the single steps, which once were elementary acts, to be intentional – and rightly so for the following reason. Before every execution of the automatism there is the molecular implementation intention to do *A*, which materially implies doing the various A_i. Though the agent can no longer recall these material implications, she has known them in detail and still has a rough idea of how the molecular action "uncoils"; by her present decision she accepts and intends not only the result but (at least) implicitly also the course to it. In the weakest case she knows that doing *A* implies many details she does not remember but after her long experience she accepts them – whatever they are. So we might say that in these cases the comprehensive implementation intention to do *A*, which is actually formed before these actions, also includes an implicit sequential implementation intention to do the various elementary actions. The latter intention makes the singular elementary behaviour intentional and hence an action, for which we are usually responsible.

5.4. Automatic Behaviour, Type 5: Habits Acquired Naturally by Independent Repetition

The last way considered here of unintentionally acquiring habits is again a form of automatic automatisation. However, this time the singular actions whose performance is habitualised are isolated; they are neither part of a more comprehensive plan, nor is there a respective general intention and still less an intention to habitualise them. The habitualisation is entirely *natural* in the sense of not being intended or sustained by any overarching intention [Shiffrin & Schneider 1977]. This implies, in particular, that no intention is part of that situation which later on will be the triggering stimulus. Therefore, the naturally automated actions can be triggered without any intention or even awareness. This mechanism is at the basis of many habits like pushing one's hair up, feeling one's chin or scratching one's head when nervous, hawking to clear one's throat (or to make a small pause), wiping away fluffs, biting one's finger nails under stress etc. Even a major component of psychic addiction, which also can complement and support physical addiction, accrues via this mechanism [Neal & Wood 2008]. As the list suggests, plenty of bad habits are also acquired in this way.

Is habitual behaviour naturally acquired by independent repetition intentional? Is it an action? It is not because there is no intention which covers the later automatic behaviour and no intention to habitualise this behaviour. Nonetheless, one might try to ascribe intentionality to them with the following general induction plus approval argument – which however fails: Repeatedly and consistently forming and executing singular intentions to do *A* in situations *C* provides the content of a hypothetic or implicit general intention; furthermore, not opposing the beginning and then successful automation may count as a weak form of implicit approval of the hypothetic general intention; therefore, we may assume a kind of generalising induction from the singular intentions to the implicit general intention. However, this proposal already suffers from an intensionality problem of the ascribed intention. If the subject in situations of type *C* always intentionally *A*s, the behaviour's description as "*A*" comes from the subject, however the situation's description as "*C*" is ours; and it is not necessarily the case that the subject has any overarching description of the situations *C* in which she *A*s. Furthermore, the general induction argument does not meet the conditions of the intentional-causalist definition of 'action', according to which in case of an action the intention *causes* the behaviour: An implicit or hypothetical intention cannot cause anything.

The upshot of this discussion then is that occurrences of habitual behaviour on the basis of naturally acquired habits are not actions and are not intentional. However, if we can suppress naturally acquired habits – and above (beginning of sect. 5) we have seen that this possibility does exist under certain conditions – then we can be responsible for them as well as for the resulting habitual behaviour (though neither the habitualisation nor the habitual behaviour in these cases is an action) – as we are responsible for many intentional and unintentional omissions, like driving with bald tires, not rendering assistance or not paying taxes.

6. Automatic Behaviour, Type 6: Mimicry

Mimicry, in its psychological meaning, is the phenomenon that humans imitate other persons unconsciously and automatically. Of course, sometimes humans also imitate other people

consciously on the basis of a proximal intention; but this is not mimicry in the psychological sense, and above all it is not of primary interest in the present context. Mimicry is a pervasive phenomenon which can concern many forms of behaviour: verbal behaviour (accent, length of pauses and length of utterances, speech rate, syntax), facial expression, many forms of bodily behaviour like posture or gestures or personal mannerisms like moving one's foot or touching one's face [Chartrand & Dalton 2009, 459-461].

A particularly important function of mimicry is recognising the emotions of other persons, which works in this way: The subject recognises the other person's facial expression, imitates it (often only in a minimalist form); the proprioception of one's own facial expression then induces the corresponding but muted emotion, which then is recognised and finally categorised [cf. Walbott 1991; Adolphs et al. 2000; Neal & Chartrand 2011]. A similar mechanism helps us understand the intention of other persons' actions by imitating their movements. So, recognising other persons' inner states is one important function of mimicry; another is to produce "social glue" in various ways [Chartrand & Dalton 2009, 462]: Perceiving oneself to be mimicked by another person leads to a feeling of like-mindedness, of similarity to and of being understood by that person and, therefore, also to liking this person more and to establishing a better relation with her as well as to being more readily convinced by her arguments because she is more easily considered to be trustworthy [Chartrand & Bargh 1999; Chartrand & Dalton 2009, 467]. A third main function of mimicry, primarily for children, is to learn certain forms of behaviour and skills by imitating them.

What is interesting here about mimicry is that mostly it is automatic [Dimberg et al. 2000; Chartrand & Bargh 1999]. On the other hand, mimicry is not automatic in the stronger sense of "rigid" and "stereotypic"; it is modulated by many factors. (High self-monitors mimic much more than low self-monitors [Cheng & Chartrand 2003]; people looking for affiliation mimic more than those who do not [ibid.]; experimental subjects in a "worker" position mimic more than those in a "leader" role [ibid.]; persons self-reporting much perspective taking of others mimic more than those with little perspective taking [Chartrand & Bargh 1999].)

The functionality of mimicry invites two main interpretations. Mimicry might be caused by hard-wired brain processes brought about by biological evolution, or it might be the (intentional or unintentional) result of intentional actions. Accordingly, in order to be able to judge about the possible intentionality of automatic mimicry we need to know the mechanism through which mimicry is brought about. There are three possible explanations, where the third explanation may be combined with the second.

1. Physiologically fixed wiring: Chartrand & Bargh [1999; reproposal: Chartrand & Dalton 2009] have proposed a hard-wired physiological "perception-behaviour link", which they concretise by referring to mirror neurons, the ideomotor theory of action and various other empirical results. According to the theory, sharing of brain resources by perception processing and by behaviour generation makes it likely that if a person perceives the execution of a certain behaviour she behaves in the same way [Chartrand & Dalton 2009, 472-474]. – However, the perception-behaviour-link theory has several and in part fatal defects. First, the physiological mechanism is, cautiously formulated, only sketched but by no means specified. Second, mimicry of

visually perceived behaviour includes several switches of the representation mode: The visually perceived behaviour must be matched with a visual representation of one's own analogous behaviour, which in turn has to be "translated" into a motor or proprioceptive representation of one's behaviour – presumably located in the mirror neurons of one's premotoric fields –; only then can some sort of motoric go signal be released. These switches are usually held to be learned ontogenetically in the course of the first year by respective experiences of co-variation. How can a theory of a physiologically hard-wired perception-behaviour link explain these facts? More generally, there is evidence that mirror neurons are modulated by experience (there is e.g. more mirror activation in pianists than in non-pianists during observation of piano playing; there are even counter-mirror neurons, which are the result of observing one action and executing a different action – e.g. I grasp an object while you release it) – which is hardly compatible with the hypothesis of an inborn fixed wiring [Heyes 2010, 579-580]. Third, as the reported results about the modulation of mimicry by present intentions, character traits, situation perception etc. show, there is at least a lot of interference of psychological processes in the mimicry process, which again is hardly compatible with the physiological perception-behaviour hypothesis.

2. Initial mimicry disposition: There might be an inborn mimicry disposition which loses its force and eventually may fade out entirely when children advance in age and are increasingly able to mimic intentionally. The disposition may be weak, intentionally suppressible, and its execution may depend on learning how to imitate other persons. Acting out the disposition would not be an action because it is not caused by an intention with a representation of the behaviour to be executed etc., instead it would be more like the working of e.g. the sucking reflex. Of course, this is a very speculative hypothesis. However, something like this is necessary to fill the explanatory gap before the beginning of more sophisticated, intentional forms of mimicry.

3. Instrumental mimicry plus habitualisation: The initial-mimicry-disposition hypothesis can explain only early but not later, much more modulated mimicry. It must be complemented by an explanation covering the latter manifestations of mimicry. In recent times mimicry mechanisms have usually been studied in adults, where the mechanisms are mostly automatised and ossified. To have a look at children's and teenagers' mimetic behaviour may be a better access for understanding the origin of these mechanisms. In this developmental period, however, there is not only much more mimicry than later, lots of it, obviously, is also driven by conscious intentions. Toddlers and kindergartners imitate demonstrated behaviour to learn it; they copy adult behaviour to feel grown-up; preschoolers precisely mimic interesting behaviour to know how it feels; elementary school children begin to imitate their peers because they want to be like them; pubescent children strengthen this tendency and try to establish rather intimate and exclusive friendships by imitating each other – thereby they learn the skills to gain sympathy by mimicry etc. There is so much intentional mimicry; much of it will be habitualised via one of the four previously discussed types of habitualisation. Much of it will be counteracted or given up later because it is childish, no longer fashionable etc. However, some habitualised forms of mimicry will remain, in particular those which turned out to be instrumentally valuable even for adults.

So, my hypothesis on how to explain automatic mimicry is a combination of the last two

mechanisms: Automatic mimicry, in early childhood, results from the initial mimicry disposition and, later, from instrumental mimicry plus habitualisation. The answer to the question of whether automatic mimicry in adults, which originates only in the second mechanism, is intentional or agential and whether we are responsible for it, then is already implied by the above section on habits (sect. 5): Intentionally learned mimicry, mimicry acquired by repetition under a general intention and routine mimicry within bigger actions are intentional and actions, and, under normal conditions, we usually are responsible for them; however, mimicry naturally acquired by independent repetition is not intentional and not an action, though we often are responsible for it.

7. Automatic Behaviour, Type 7: Semantically Induced Automatic Behaviour

Bargh has proposed still another mechanism of automatic behaviour: automatic action via *auto-motives*: A "goal" is activated unconsciously not by an "act of will" but by situational features, e.g. priming; subsequently the "goal" produces the same effect as if it were put into motion intentionally; in several experiments none of the subjects reported to have pursued the goal hypothesised by the experimenter. Some experimental examples are: When primed (even subliminally) by words from the lexical field of 'achievement' subjects produced better results in a verbal task or they cheated in order to improve their scores. [Bargh 1989; 1990; Bargh & Chartrand 1999, 469-473.] With "goal" Bargh, unusually, means a certain mental state. According to one interpretation, this state might be a goal intention; then, however, the described mechanism would not produce automatic behaviour in the above sense (condition 4 would be violated). Since Bargh distinguishes between a "goal" and an "act of will" he seems to mean something else, which I would interpret as a *non-volitive executive mental state*, i.e. a state that represents a goal-directed behaviour and usually causes this behaviour in an organised way (hence it is an executive mental state) but it lacks the volitive dimension of intentions and hence is not an intention because it has not developed from desires, so that no optimality belief about the executed behaviour corresponds to it [cf. Lumer 2005; 2013]. A particular feature of Bargh's examples is that the executive state is activated by semantic association. Therefore, I call the mechanism under discussion "*semantically induced automatic behaviour*" and its key state "*semantically induced executive state*".

These conceptual differentiations also help to systemise the various kinds of automatic behaviour. Much automatic behaviour is caused in an organised way by non-volitive executive states. The executive state is activated in various ways: 1. In actions caused by distal implementation intentions, as a consequence of a respective programming by the distal intention, perceiving a specific stimulus activates the executive state. 2.-5. In habitual behaviour as a consequence of a respective programming by (intended or unintended) habitualisation, perceiving a specific stimulus activates the executive state. (However, the final stage of habitualisation may imply shortcutting the process of action production to such a degree that even executive states in the form of unconscious goal representations are skipped.) 6. In automatic natural mimicry, as a consequence of an inborn mimicry disposition, perceiving the action to be mimicked activates the executive state. 7. Finally, in semantically induced automatic behaviour, semantic inputs

associatively activate a representation of the later executed behaviour, which then activates the respective executive state. Now however, it is unclear why and on which basis the behaviour representation activates the executive state – after all, prima facie, this activation is pathological (pace ideomotor theory): if every action representation caused the represented action, we could e.g. not deliberate about which course of action to take. Bargh is silent on this question. If the transition is due to arousal of a respective *desire* the resulting executive state probably is a proximal intention and the generated behaviour no longer automatic. If the transition is due to some *confusion* the reliability condition of the definition of 'automatic behaviour' is not met. Both of these possibilities are sometimes empirically real but they are not cases of automatic behaviour.

The kernel of Bargh's examples which might be semantically induced automatic behaviour consists of unconsciously *modulating* choices of actions to be taken on independent grounds; e.g. fulfilling a verbal task is modulated to be done with more effort. So far there is no example where a fresh behaviour is automatically started from scratch by semantic induction – which would be the only way to prove that behaviour can be automatically induced without mediation by proximal intentions.

Altogether then, semantically induced automatic behaviour so far is only a hypothesis, which is still explanatorily incomplete (how does the behaviour representation activate the executive state?) and insufficiently sustained by experiments. However, it has not been falsified either and not yet studied sufficiently, such that it remains an interesting hypothesis. Because of the explanatory gap, unfortunately, there is no basis to decide whether the step from behaviour representation to an executive state might rely on some (earlier) intention, which could make semantically induced automatic behaviour agential, or whether we can intentionally prevent this step, such that we might be responsible for not preventing it and indirectly responsible for the automatic behaviour itself.

8. Other Philosophical Approaches to Automatic Actions

The broad survey of various forms of automatic behaviour has to stop here. Before considering the consequences for the intentionality of automatic behaviour and our responsibility for it, a brief discussion of some competing philosophical explanations and assessments of automatic behaviour may elucidate some merits of the analysis provided here.

Perhaps the best-known approach, which also tries to explain automatic behaviour, is Gendler's conception of "aliefs" [Gendler 2008,]. Aliefs, according to Gendler, are fundamental (i.e. irreducible) mental states with a representational, affective and conative content and function [ibid. 641; 643-644]; someone walking on a "skywalk" with a glass floor over a chasm may e.g. alieve: 'Really high up, long long way down. Not a safe place to be! Get off!!' [ibid. 635]. Aliefs are associative, automatic and arational; and they may be in contrast to our conscious beliefs – as in the example just given the belief that the "skywalk" is safe –, and thus lead to otherwise unexplainable behaviour [ibid. 641; 642; 646]. Gendler proposes to explain also automaticity in action with the help of aliefs, especially experimental examples provided by Bargh [e.g. Bargh et

al. 1996], such as priming subjects with the concept of 'politeness', which, unnoticed by the subjects, leads to more polite behaviour [ibid. 656-661]. Now, apart from the fundamental doubt whether aliefs exist at all, whether they are not reducible e.g. to causally connected primitive beliefs, emotion inducing valuations and desires, the more specific concern in the present context is that all examples of automatisms brought forward by Gendler include priming before a conscious decision which causes a somewhat modified but consciously triggered behaviour. These examples may involve automatic processes, but because the action is caused by a conscious proximal intention, according to the above definition, they do not instantiate automatic behaviour. Perhaps the set of Gendler's examples is chosen too narrowly, and the conception of 'alief' by itself has the potential to serve to explain also automatic behaviour. However, there are three strong reasons to doubt also this hypothesis. First, 'automatic behaviour' has been defined above in a way which excludes involvements of many kinds of mental events (conscious triggering, causation by a proximal intention, conscious guidance during execution). Aliefs, however, are defined as "thick" mental states with emotional and conative components, whose intervention would probably lead to at least conscious traces and / or intervention on one of the excluded processes. Second, the conative component of aliefs is comparable to desires about certain *outcomes* of actions but not about situation-specific actions themselves (perhaps apart from negative actions, i.e. omissions). Hence for causing a (positive) action they need to be brought together with representations of presently possible actions and perhaps considerations of other aspects of these actions. Even if this kind of elaboration were unconscious it would probably involve a newly formed (unconscious) proximal intention, causing an unconsciously deliberated action, which, however, by definition is not automatic behaviour. Third, in any case, the types of automatic behaviour analysed above (sections 4-7) cannot be explained by aliefs; they are too automatic for permitting the intervention of a further and so massive mental state involving affective and desiderative components.

Roughley has proposed Bargh's auto-motives as a general solution to the problem of automatic behaviour for intentional-causalist theories of action [Roughley 2007]: The intentions are there, they are only unconscious. Problems with this proposal are, however: First, the mechanism of semantically induced automatic behaviour, according to what has been said in the last section, so far is only hypothetical; hence the proposal is based more on a hope than on empirical evidence. Second, even if semantically induced automatic behaviour (Bargh's auto-motives) exists it would (also according to Bargh himself) be only one of several mechanisms of automatic behaviour; and the challenge for intentional causalism would be unanswered for the quantitatively far more important set of automatic behaviour. Third, after the analysis in the last section, Bargh's auto-motives are (semantically induced) mere executive states, which, however, according to the more demanding requirements of intentional causalism (cf. above, sect. 1), are not yet intentions. Therefore, Roughley's proposal would not meet the challenge for intentional causalism even for this small group of automatic behaviour.

9. Intentionality of Automatic Actions and Responsibility for Them

Let us take stock. There may be some further types of automatic behaviour, but the quantitatively most important probably were included in the above enquiry. The results regarding the intentionality and the agential character of the different forms of automatic behaviour were these. Most automatic behaviours, namely actions caused by distal implementation intentions (type 1), intentionally learned routine actions (type 2), habitual actions acquired by repetition under a general intention (type 3) and routine actions within molecular actions (type 4) as well as the mimicry versions of the latter three forms of habitual actions (types 6.2-6.4), have been found to be intentional and actions, which usually implies the agent's responsibility for them. Habitual behaviours acquired naturally by independent repetition (type 5) as well as the mimicry version of them (type 6.5), however, are neither intentional nor actions, though the agent is, nonetheless, often responsible for them (e.g. if the agent is able to suppress them and socially required to do this). Only behaviour emerging from the initial mimicry disposition (type 6.1) is neither intentional nor an action and we are not responsible for it; however, this type of behaviour already fades out during childhood. Finally, semantically induced automatic behaviour (type 7) is still too unclear (whether it exists and how it works) for assessing whether it is agential or whether we can be responsible for it.

With these explanations also the examples introduced above for illustrating the challenge by automatic actions can easily be explained such that the challenge dissolves: The *fare dodger* and the *obese* act habitually, and their habit originally is based on their intention such that their habitual actions are intentional and they are responsible for it. The *reckless driver* is right in saying that she could not act differently in the specific situation. But she could have and was required much earlier to intentionally habitualize a more considerate style of driving. Because she omitted this she is indirectly responsible for her reckless driving and for the accident. The *mafioso*, finally, did not have a proximal intention to press the trigger for blowing up his enemy. But he had a distal intention to do just this (the last time occurrent when the accomplice tells him that the enemy is approaching). And because this distal intention causes the respective behaviour in the right way this behaviour is an intentional action, and the mafioso is responsible for it.

The criterion for the just repeated action and intentionality attributions was intentional causalism: a behaviour is intentional if it is caused in a non-deviant way by a corresponding intention. Here is not the place to discuss this criterion. However, there are some intentionality attributions which even a proponent of intentional causalism may find problematic. One is this: Since after habitualisation, habitual actions function independently of the respective intention, for some time they may be executed automatically even against a newly formed opposite intention. Nonetheless, according to the criterion used here, this automatic behaviour would count as intentional because of the original intention (if the habit was acquired via intentional habitualisation or under a general intention). This seems to be odd.[7] The appearance of oddity in such cases,

7 Bruno Verbeek raised this objection in a personal conversation.

however, is not due specifically to automaticity but, more generally, due to, first, the (long) interval between the intention and its intended actional or further consequences, which allows the agent to change her mind in the meantime, and, second, to the ballistics of the process leading from the intention to the originally intended consequences, which after a certain point forestalls to retract them and to block their realisation. There are other (i.e. without automaticity) situations with the same features – someone has signed an irrevocable long-term contract or posted a letter, shot an arrow or said something in a media conference, which she now, even before the planned effects (i.e. the contract conditions to be fulfilled, the letter arriving at its destination, the arrow striking the target, the greater public hearing the message) occur, finds problematic. With respect to such situations we do not hesitate to consider the action or its planned consequences to be intentional though the agent has now changed the action's evaluation and would like to retract it – e.g. because she now has learned about originally unforeseen further consequences (which, of course, are not intentional). The systematic reason for this, prima facie paradoxical, intentionality attribution is again intentional causalism and hence secunda facie quite plausible: the later, planned event (unlike the originally unforeseen consequences) is exactly what the agent intended and was caused by the intention in the right way. The only problem is that the intention is more powerful than the agent likes in this situation – but only in this situation. The same reasoning should hold for habitualised automatic actions as consequences of intentional habitualisation. However, one lesson to be learned from such cases – like in legal matters – is: be careful with habitualisations because of their powerful dynamics.

That we have no occurrent intention when starting or executing the respective behaviour is a defining feature of automatic action, but often we do not even know this intention if we are asked for it right after the behaviour and often we even know nothing about the automatic behaviour itself. Then a further objection is: How can such automatic behaviour be an action? The reason for this ignorance mainly is that the intention which, according to the present proposal, in case of automatic *actions* is the cause of the respective behaviour was occurrent only too long ago and that to know it is of no help for executing the behaviour; hence forgetting it is a relief for our memory. Sometimes we need such knowledge in order to give an account of our action or to be able to change our habits; but in such cases mostly a rough idea of the action and its intention is sufficient to serve these purposes, and this rough idea often is still present. So, altogether this ignorance can sometimes be a – marginal – handicap but on the contrary it is usually advantageous. What is really important for the sense and essence of intentional actions is that the intention causes the respective behaviour (in the right way); this condition, however, is fulfilled by automatic actions. Au contraire, the existence of automatic actions shows that this essence can also be realised in a very powerful way with long-range control of our behaviour and without any need to further consciously care about its execution.

One might be tempted to summarise the just developed results on the responsibility for our automatic behaviour by saying that we are virtually always responsible for it. However, this would be grossly false. First, adults *are usually* responsible for their intentional actions; they *can* also be responsible for omissions and unprevented events but only under specific conditions in a limited

number of cases. Responsibility for mere automatic *behaviour* is only an *indirect responsibility* resting on the responsibility for having forborne to eliminate the automatism. In order to have the latter, direct responsibility, the agent must command effective means of elimination, which she does not always have at hand, and she must have the moral, legal or something like a prudential positive duty to eliminate the automatism. The reason for this strong restriction is this: As just has been said, we can be responsible for omissions (even unintended omissions) and for preventable but unprevented events; but the set of omissions and preventable but unprevented events is so huge that we cannot be responsible for all of them. We are responsible only for a small subset, namely those omissions and unprevented events for which a corresponding positive duty of action and prevention respectively exists. [Lumer 2012] The unprevented event in our case is the automatic behaviour, and the omission consists in not stopping the automatism. Therefore, we are *morally / legally* (indirectly) *responsible* for this automatic behaviour if we have effective and sensible means to stop it and if this behaviour is morally / legally obligatory or prohibited; and we are *prudentially* (indirectly) *responsible* for it if we have effective and sensible means to stop it and if this behaviour is particularly good and rational or rather harmful or irrational.

Second, we are mostly directly responsible for our automatic *actions* but not always. A very important group of such cases where we are not directly responsible is habits acquired in periods of reduced or non-responsibility, in particular during childhood and early adolescence. Many habits are acquired during childhood even on the basis of general intentions (though the share of naturally acquired habits at this age will be bigger than it is in adults), so that their later execution, according to the above discussion, is intentional and each time an action. Since children are not yet responsible or responsible only to a reduced degree they may not be responsible for the habitualisation. And this lack of responsibility may extend to a lack of direct responsibility for the later automatic action even of the adult. This, however, does not preclude that the adult has an *indirect* responsibility for his automatic action in the way analysed in the last paragraph. Knowledge about secondary responsibility and about effective means against undesired automatisms should prompt one to check one's automatisms for their adequacy and to perhaps change them.

The foregoing considerations show that it is important to distinguish the various automatisms and their origins in order to be able to attribute intentionality, agential character and direct or indirect responsibility.

10. Conclusion

The result of the above discussion is that most forms of automatic behaviour are intentional actions, for which we are usually responsible, and that we are also responsible for much of the remaining automatic behaviour. This is good news. Remember, intentionality in the explained strong sense guarantees that our ego, with which we identify ourselves, controls our behaviour, so that with the explanation given above, which identifies the intention behind automatic actions, even these actions in most cases are under our control. Furthermore, this ego is rational to a certain degree and so far

also sensible to reasons given by society, which in turn makes subjects responsible. Therefore, the concern that, through the discovery of its automaticity, 95% of the behaviour which has so long been taken to be actions, may prove not to be intentional, or not actions, or not something for which we can be responsible or which is done for reasons, turns out to be unfounded. Hence there is no reason, at least this concern isn't one, to abandon or to fundamentally revise the intentional-causalist concept of action or, what is more, our everyday and legal practice of ascribing responsibility and reasons to actions.

However, the survey also revealed that the mechanism of intentional agency at several points, via automatisms, is more powerful than usually assumed. That we can intentionally acquire automatic routine actions, which later relieve us from conscious control of these actions, was no real news. However the findings about the implications of the other mechanisms probably have more news value for most of us: Distal implementation intentions are a sort of self-programming for the future and can be executed independently of a reformulation as present implementation intentions; repeatedly acting under a general intention can lead to a respective habitualisation with the consequence of automatically executing these actions without further intentional approval; the same holds for repeatedly executing routine sequences in molecular actions and for often repeating the same action in the same situation. This, newly acquired, knowledge about the greater impact of our intentional actions should also be an admonition to be careful in forming intentions and executing actions which can have this additional power. And the fact that habits continue automatically without needing a further intentional approval as well as the possibility that even an intentionally acquired habit later on may no longer fit to the intentions and interests changed or developed in the meantime should be a warning to occasionally check whether one's habits are still adequate and, if the result is negative, to possibly counteract them with the help of the strategies explained above.

Acknowledgements: My thanks to two anonymous referees for their important and helpful proposals. And I would like to thank all the participants of the workshop on “Everyday reason” for the fruitful discussion of this paper, and especially Leon de Bruin and Bruno Verbeek for further valuable suggestions.

References

Adolphs, R.; Hannah Damasio; D. Tranel; G. Cooper; Antonio R. Damasio (2000): A Role for Somatosensory Cortices in the Visual Recognition of Emotion as Revealed by Three-Dimensional Lesion Mapping. In: Journal of Neuroscience 20. Pp. 2683-2690.

Babcock Gove, Philip (1993) (Ed.): Webster's Third New International Dictionary of the English Language Unabridged. Editor in Chief Philip Babcock Gove. Springfield, MA: Merriam-Webster. 120a; 2663 pp.

Bargh, John A. (1989): Conditional Automaticity. Varieties of Automatic Influence in Social Perception and Cognition. In: James S. Uleman; John A. Bargh (Eds.): Unintended

Thoughts. New York; London: Guilford Press. Pp. 3-51.

Bargh, John A. (1990): Auto-Motives. Preconscious Determinants of Social Interaction. In: E. T. Higgins; R. M. Sorrentino (Eds.): Handbook of Motivation and Cognition. New York; London: Guilford Press. Pp. 93-130.

Bargh, John A. (2005): Bypassing the will. Toward demystifying the nonconscious control of social behavior. In: Ran R. Hassin; James S. Uleman; John A. Bargh (Eds.): The new unconscious. New York: Oxford U. P. Pp. 37-58.

Bargh, John A. (2008): Free will is un-natural. In: J. Baer; J. C. Kaufman; R. F. Baumeister (Eds.): Are we free? Psychology and free will. Oxford; New York: Oxford U.P. Pp. 128-154.

Bargh, John A.; Tanya L. Chartrand (1999): The Unbearable Automaticity of Being. In: American Psychologist 54. Pp. 462-479.

Bargh, John A.; M. Chen; L. Burrows (1996): Automaticity of social behavior. Direct effects of trait construct and stereotype activation on action. In: Journal of Personality and Social Psychology 71. Pp. 230-244.

Baumeister, Roy F.; Ellen Bratslavsky; Mark Muraven; Dianne M. Tice (1998): Ego depletion. Is the active self a limited resource? In: Journal of Personality and Social Psychology 74. Pp. 1252-1265.

Baumeister, Roy F.; K. L. Sommer (1997): Consciousness, free choice, and automaticity. In: R. S. Wyer, Jr. (Ed.): Advances in social cognition. Vol. X. Mahwah, NJ: Erlbaum. Pp. 75-81.

Brandstätter, Veronika; Angelika Lengfelder; Peter M. Gollwitzer (2001): Implementation Intentions and Efficient Action Initiation. In: Journal of Personality and Social Psychology 81. Pp. 546-560.

Chartrand, Tanya L.; John A. Bargh (1999): The chameleon effect. The perception–behavior link and social interaction. In: Journal of Personality and Social Psychology 76. Pp. 893-910.

Chartrand, Tanya L.; Amy N. Dalton (2009): Mimicry: Its ubiquity, importance, and functionality. In: Ezequiel Morsella; John A. Bargh; Peter M. Gollwitzer (Eds.): Oxford Handbook of Human Action. Oxford; New York: Oxford U.P. Pp. 458-483.

Cheng, Clara Michelle; Tanya L. Chartrand (2003): Self-monitoring without awareness. Using mimicry as a nonconscious affiliation strategy. In: Journal of Personality and Social Psychology 85. Pp. 1170-1179.

Dimberg, Ulf; Monika Thunberg; Kurt Elmehed (2000): Unconscious facial reactions to emotional facial expressions. In: Psychological Science 11. Pp. 86-89.

Gendler, Tamar Szabó (2008): Alief and Belief. In: Journal of Philosophy 105. Pp. 634-663.

Gollwitzer, Peter M. (1999): Implementation Intentions. Strong Effects of Simple Plans. In: American Psychologist 54. Pp. 493-503.

Graybiel, Ann M. (2008): Habits, rituals, and the evaluative brain. In: Annual Review of Neuroscience 31. Pp. 359-387.

Heckhausen, Heinz (1987): Intentionsgeleitetes Handeln und seine Fehler. [Intention-guided action and ist slips.] In: Heinz Heckhausen; Peter M. Gollwitzer; Franz E. Weinert (Eds.): Jenseits des Rubikon. Der Wille in den Humanwissenschaften. Berlin [etc.]: Springer. Pp. 143-175.

Heyes, Cecilia (2010): Where do mirror neurons come from? In: Neuroscience and Biobehavioral Reviews 34. Pp. 575–583.

Hieronymi, Pamela (2009): Two Kinds of Agency. In: Lucy O'Brien; Matthew Soteriou (Eds.): Mental Actions. Oxford: Oxford University Press. Pp. 138-162.

James, William (<1890> 1950): The Principles of Psychology. (1890.) 2 Vols. Reprint: NewYork: Dover 1950. xii; 689; vi; 702 pp.

Kenny, Anthony (1978): Freewill and Responsibility. London; New York: Routledge & Kegan Paul 1978. viii; 101 pp.

Kihlstrom, John F. (1987): The Cognitive Unconscious. In: Science 237. Pp. 1445-1452.

Koch, Christof; F. Crick (2001): The zombie within. In: Nature 411,6840. P. 893.

Libet, Benjamin (1985): Unconscious cerebral initiative and the role of conscious will in voluntary action. In: Behavioral and Brain Science 8. Pp. 529-566.

Lumer, Christoph (2005): Intentions Are Optimality Beliefs – but Optimizing what? In: Erkenntnis 62. Pp. 235-262.

Lumer, Christoph (2007): An Empirical Theory of Practical Reasons and its Use for Practical Philosophy. In: Christoph Lumer; Sandro Nannini (Eds.): Intentionality, Deliberation and Autonomy. The Action-Theoretic Basis of Practical Philosophy. Aldershot: Ashgate 2007. Pp. 157-186.

Lumer, Christoph (2008): Abwegige Absichtsrealisierung und Handlungssteuerung. Eine intentional-kausalistische Erklärung. [Deviant realisation of intentions and action control. An intentional-causalist explanation.] In: Internationale Zeitschrift für Philosophie 1/2008. Pp. 9-37.

Lumer, Christoph (2012): Attributive Verantwortung – eine Theorieskizze. [Attributive responsibility – Outline of a theory.] In: Oliver Petersen; Dagmar Borchers; Thomas Spitzley; Manfred Stöckler (Eds.): Proceedings von GAP.7 Nachdenken und Vordenken – Herausforderungen an die Philosophie. University Duisburg-Essen (DuEPublico) 2012. Pp. 703-722. Web-publication, URL: <http://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-29983/Proceeding_GAP7_Nachdenken_Vordenken.pdf>, 10.8.12.

Lumer, Christoph (2013): The Volitive and the Executive Function of Intentions. In: Philosophical Studies 166. Pp. 511-527.

Lumer, Christoph (2014a): Libet's Experiments and the Possibility of Free Conscious Decision. In: Christoph Lumer (Ed.): Morality in Times of Naturalising the Mind. Boston; Berlin: de Gruyter 2014. Pp. 63-103.

Lumer, Christoph (2014b): The Effectiveness of Intentions – A Critique of Wegner. In: Christoph Lumer (Ed.): Morality in Times of Naturalising the Mind. Boston; Berlin: de Gruyter 2014. Pp. 105-124.

Lumer, Christoph (under review): Unconscious Motives and Actions – Responsibility and Agency.

Mele, Alfred R. (2009): Mental Action. A Case Study. In: Lucy O'Brien; Matthew Soteriou (Eds.): Mental Actions. Oxford: Oxford University Press. Pp. 17-38.

Neal, David T.; Tanya L. Chartrand (2011): Embodied Emotion Perception. Amplifying and Dampening Facial Feedback Modulates Emotion Perception Accuracy. In: Social Psychological and Personality Science 2. Pp. 673-678.

Neal, David T.; Wendy Wood (2008): Linking addictions to everyday habits and plans. In: Behavioral and Brain Sciences 31. Pp. 455-456.

Neal, David T.; Wendy Wood; Mengju Wu; David Kurlander (2011): The pull of the past. When do habits persist despite conflict with motives? In: Personality and Social Psychology Bulletin 37. Pp. 1428-1437.

Neumann, Odmar (1984): Automatic Processing. A Review of Recent Findings and a Plea for an Old Theory. In: W[olfgang] Prinz; A[ndries] F[rans] Sanders (Eds.): Cognition and Motor Processes. Berlin [etc.]: Springer. Pp. 255-293.

Neumann, Odmar; Wolfgang Prinz (1987): Kognitive Antezedentien von Willkürhandlungen. [Cognitive antecedents of voluntary actions.] In: Heinz Heckhausen; Peter M. Gollwitzer; Franz E. Weinert (Eds.): Jenseits des Rubikon. Der Wille in den Humanwissenschaften. Berlin [etc.]: Springer. Pp. 195-215.

Norman, Donald A. (1981): Categorization of Action Slips. In: Psychological Review. 88. Pp. 1-15.

Posner, Michael I.; Charles R. R. Snyder (1975): Attention and cognitive control. In: Robert L. Solso (Ed.): Information processing and cognition. The Loyola symposium. Hillsdale, NJ: Erlbaum. Pp. 55-85.

Quinn, Jeffrey M.; Anthony T. Pascoe; Wendy Wood; David T. Neal (2010): Can't control yourself? Monitor those bad habits. In: Personality and Social Psychology Bulletin 36. Pp. 499-511.

Reason, James (1979): Actions Not as Planned. The Price of Automatization. In: Geoffrey Underwood; Robin Stevens (Eds.): Aspects of Consciousness. Vol. 1.: Psychological Issues. London [etc.]: Academic Press. Pp. 67-89.

Roth, Gerhard (2001): Fühlen, Denken, Handeln: Wie das Gehirn unser Verhalten steuert [Feeling, thinking, acting: How the brain controls our behavior]. Frankfurt, Main: Suhrkamp.

Roughley, Neil (2007): Hilberts Krawatte, Ryles Clown und Gehlens Schlüssel. Zur Analyse von Gewohnheitshandlungen. [Hilbert's tie, Ryle's clown, and Gehlen's key. An analysis of habitual action.] In: Zeitschrift für philosophische Forschung 61. Pp. 188-206.

Shiffrin, Richard M.; S. T. Dumais (1981): The development of automatism. In: J. R. Anderson (Ed.): Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum. Pp. 111-140.

Shiffrin, R. M.; W. Schneider (1977): Controlled and automatic human information processing. II. Perceptual learning, automatic attending, and a general theory. In: Psychological Review 84. Pp. 127-190.

Strawson, Galen (2003): Mental Ballistics. The Involuntariness of Spontaneity. In: Proceedings of the Aristotelian Society 103. Pp. 227-256. – Reprinted in: Idem: Real Materialism and Other Essays. Oxford: Clarendon Press 2008. Pp. 233-253.

Vargas, Manuel R. (2013). Building Better Beings: A Theory of Moral Responsibility. Oxford: Oxford U.P. x; 345 pp.

Vierkant, Tillmann (2013): Managerial control and free mental agency. In: Andy Clark; Julian Kiverstein; Tillmann Vierkant (Eds.): Decomposing the Will. Oxford: Oxford U.P. Pp. 283-297.

Walbott**,** H. G. (1991): Recognition of emotion from facial expression via imitation? Some indirect evidence for an old theory. In: British Journal of Social Psychology 30. Pp. 207-219.

Wegner, Daniel M. (2002): The Illusion of Conscious Will. Cambridge, Mass.; London: MIT Press. xi; 405 pp.

Wilson, Timothy D. (2002): Strangers to Ourselves. Discovering the Adaptive Unconscious. Cambridge, MA; London: Belknap Press of Harvard University Press. viii; 262 pp.

Wood, Wendy; David T. Neal (2007): A new look at habits and the habit-goal interface. In: Psychological Review 114. Pp. 843-863.

Wu, Wayne (2013): Mental Action and the Threat of Automaticity: In: Andy Clark; Julian Kiverstein; Tillmann Vierkant (Eds.): Decomposing the Will. Oxford: Oxford U.P. Pp. 244-261.