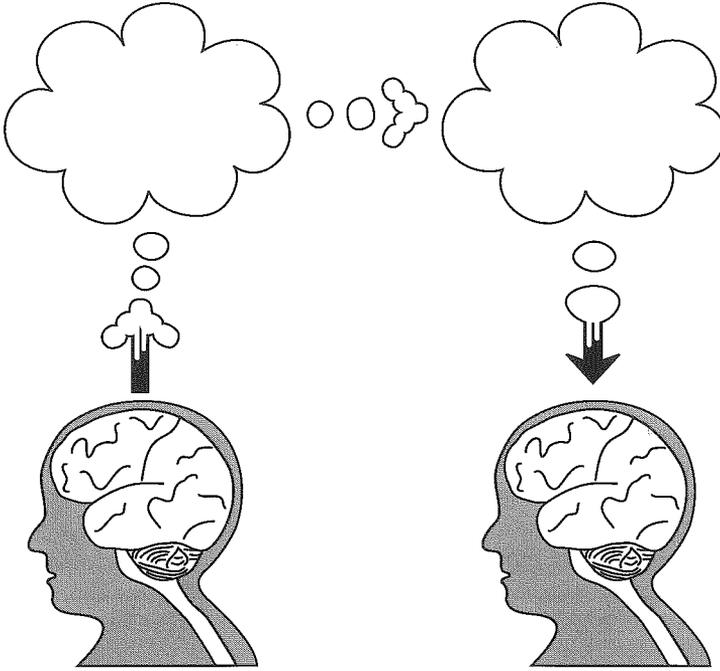


actions, therefore he regarded the pineal gland as a promising place for a channel between brain and soul. First, the pineal is located in the centre of the brain, which makes it ideal as the headquarters for controlling sensorimotor functions. Second, the pineal is unified, forming a single nucleus in the brain, unlike many others that exist in duplicate, one in each cerebral hemisphere. The anatomical unity of the pineal in the brain corresponds nicely with the mental unity of consciousness. Third, the pineal is located at the roof of the third brain ventricle. In medieval theories that reigned up until the times of Descartes, the soul was thought to be located in the fluids (in the form of “animal spirits”) that fill up the brain ventricles, not in the neural tissues at all. In his theory, Descartes preserved the ancient idea that the fluids in the brain ventricles mediate the messages between the body, the brain and the soul. Vibrations of the pineal were believed to be transformed into vibrations of the fluids in the third ventricle, which were then conducted to the muscles as commands to move the extremities. Nerves were believed to be hollow tubes, like hydraulic water-pipes, that mediate distant causal effects to the muscles through changes of hydraulic pressure in the nerves. Stimulation of the sensory nerves, respectively, was believed to be conducted to the brain, to be transformed into vibrations of the fluids in the ventricles and thereby communicated to the pineal, which forwarded them to the soul.

This is Descartes’ interactionistic dualism in a nutshell (see Figure 1.1). Although nearly 400 years has passed since he put forward these ideas, his theory continues to be debated and criticized in the recent literature on consciousness. Of course, there are very few scientists or philosophers today who defend something like interactionist dualism. Yet, no-one has been able to solve the problem of consciousness in terms of a nondualistic theory either, so Descartes is not any worse off than anyone else! We still do not have a scientific theory that would explain, once and for all, how the subjective psychological reality is connected with the objective physical and biological reality. Hence, dualism has not been defeated for good yet. It just *might* make a comeback. But if the nonphysical soul would make a comeback and be seriously considered as a scientific hypothesis, then we would have to revise our current scientific world-view quite a bit! Most scientists and philosophers, however, are not prepared to do that, thus they will try almost anything else first before they would consider a dualistic theory of consciousness. They also see no evidence whatsoever supporting dualism. Thereby dualism has become the very last resort, only acceptable if nothing else works.

### **Epiphenomenalism**

Perhaps the gravest problem for interactive dualism is to explain *mental causation* or how the nonphysical soul-stuff causally influences physical brain activity so that our behaviour is guided by the mind. Epiphenomenalism represents a way to get rid of that problem: it denies the possibility of mental causation, the idea that events in consciousness have effects in the objective physical world. In other words, the non-physical mental reality cannot causally influence physical matter or brain activity. However, epiphenomenalism does allow causation the other way round: from the physical realm to the mental realm. Physical changes in the sensory systems and in the brain cause conscious events in our subjective psychological reality. Thanks to the unidirectional causation from external world to brain to consciousness, we consciously

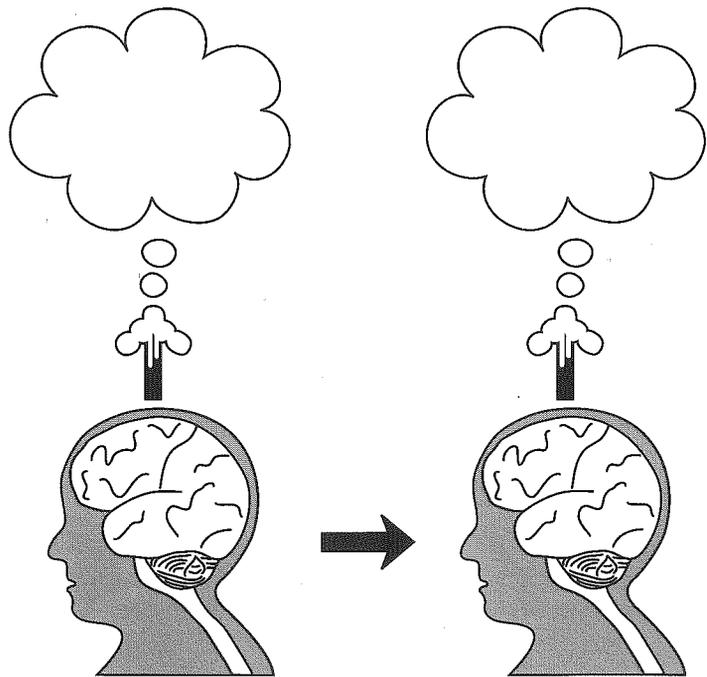


**Figure 1.1** Interactionist dualism

The figure depicts how brain and mind are in two-way causal interaction through time. First, brain activity has causal effects on consciousness (brain  $\rightarrow$  consciousness causation, symbolized by the arrow pointing up). This happens for example when visual information arrives in the visual cortex and causes a conscious visual percept in the mind of the observer. Second, contents of consciousness within the same person's mind have causal effects on each other (consciousness  $\rightarrow$  consciousness causation, symbolized by the arrow between the two clouds of consciousness). This happens for example when a conscious percept causes a conscious thought or an intention to act. Third, at least some contents of consciousness have causal effects on brain activity and thereby also on behaviour (consciousness  $\rightarrow$  brain causation (or *mental causation*), symbolized by the arrow pointing down). This happens for example when a conscious intention to act causes changes in the motor cortex where behaviour is initiated

sense and perceive the world around us. Physical brain activity causes two types of things: further physical changes in the brain, which ultimately cause all our observable behaviour; and events in consciousness, such as sensations, percepts, thoughts, intentions, action plans. But the nonphysical events in consciousness have no causal powers whatsoever. They do not cause any changes in the brain nor do they cause any further conscious events in the mind. Conscious events are, causally, a dead end. They have been compared to immaterial shadows cast by physical brain activity, shadows that just hang around, or perhaps "above" brain activity, but do absolutely no work whatsoever and have no effects on anything (see Figure 1.2).

Thus, the strength of epiphenomenalism is that it can explain all human behaviour by purely physical causation. Physical stimulation causes physical brain states, which cause further physical brain states, which cause our observable



**Figure 1.2** Epiphenomenalism

The figure depicts how brain and mind are only in one-way causal interaction through time. First, brain activity has causal effects on consciousness (brain > consciousness causation, symbolized by the arrow pointing up). This happens for example when visual information arrives in the visual cortex and causes a conscious visual percept in the mind of the observer. Second, the initial brain activity in the visual cortex causes further activities that spread to other cortical areas (purely physical causation between brain activities is symbolized by the black arrow pointing from left to right). The further brain activities have simultaneously further causal effects on consciousness (the second arrow pointing up). Consciousness itself has no causal effects on anything (there are no causal arrows originating from consciousness)

behaviour. And that is the whole story there is to tell in terms of scientific explanation. Objective causal mechanisms in the physical world, such as neuronal firing patterns in the brain, are sufficient to explain all aspects of human behaviour. As scientific explanations are basically nothing but detailed descriptions of the causal mechanisms that determine the behaviour of observable phenomena, we are blessed with a science of psychology (or cognitive neuroscience) that explains all the objective data we will ever have, by simply referring to fully physical (chemical, biological, neural, etc.) mechanisms. The nonphysical (mental, subjective, conscious) phenomena are not needed in the explanation of anything. They are *epi*-phenomena, mere secondary effects or side-effects of brain activity.

Unfortunately, getting rid of mental causation is a double-edged sword. Undoubtedly, the greatest weakness of epiphenomenalism is precisely the fact that it gives our mental life *no* active role at all in the world. Our streams of consciousness

are mere shadows dragged along by the brain, and we ourselves as conscious beings are nothing more than passive onlookers whose existence makes no difference whatsoever to anything in the world. We are at best in for a joyride, helplessly floating in the stream of consciousness but under the illusion of having control of our behaviour or of having a free will to decide what to do next. In actual fact, we have no control whatsoever over anything and our conscious will is but a deception. We are like shadows that confuse themselves with the physical things whose shadows we are, believing that we are in control because the physical things change in perfect correlation with our experiences. The shadow cast by a hammer might believe that it is in control of the hammer's behaviour and causes the physical hammer to move and hit the nail and the physical nail to sink in. We make exactly the same mistake in believing that we, as conscious subjects, are in control of our own behaviour. The physical causes in the brain do all the work and our minds follow the physical causes like shadows.

The fatal problem here is that the view of ourselves as mere passive shadow-like observers is in stark conflict with our beliefs and direct experiences about ourselves. According to epiphenomenalism, all of the following statements are false: the feeling of thirst causes me to drink; the feeling of pain causes me to take an aspirin; my deliberate plans and careful considerations cause me to take one course of action rather than another; what I visually experience to be around me causally guides my behaviour; and so on. Our everyday experience about being in control, about our subjective psychological reality causally influencing our behaviour, is simply so strong that we are not prepared to give it up just like that. The view of ourselves as mere conscious puppets whose strings are attached to the brain of an otherwise nonconscious biomechanical zombie is very difficult to accept. On the contrary, we take ourselves as conscious human beings whose observable behaviour is largely determined by the mental events happening in our subjective psychological reality. We – our minds – *do* make a difference in the world and we *can* change the world by having our inner conscious life guide our external behaviour.

Therefore, it is hard to accept an epiphenomenalist view of ourselves. Our lives as conscious beings would be totally useless; we would be like helpless spectators locked in a giant virtual-reality movie theatre, forced to watch the movie without being able to influence its events.

Epiphenomenalism is rarely put forward as an explicit theory of consciousness. Rather, it is a position (or a trap) where many a theorist finds himself after painting himself into a corner by first putting forward a thoroughly physicalist theory of mind. The purely physicalist theory may nicely explain sensation, perception, cognition, action and behaviour as objective, nonconscious phenomena. Epiphenomenalism creeps in only after there is no explanatory role left for our inner conscious life. In the theory, all seems to work just perfectly even without an inner life. Hence, to give at least some kind of place to our subjective psychological reality in a physicalist theory, as a last resort consciousness is interpreted as an epiphenomenon that hangs somewhere above the physical, somewhere beyond all the objective neural and cognitive mechanisms that do all the real work. An epiphenomenon cannot interfere in any way with the physical workings of the “real” mechanisms, and therefore it can be easily added to the physical theory without changing anything else in the picture.

In the 1870s the British physiologist and philosopher Thomas Henry Huxley

defended a neurophysiological version of epiphenomenalism. In the 1980s, the American linguist and cognitive scientist Ray Jackendoff (1987) defended a cognitive theory of consciousness that implied epiphenomenalism, and in the 1990s the Australian philosopher David Chalmers (1996) formulated a philosophical theory of consciousness that makes subjective experience explanatorily irrelevant and therefore his view at least approximates epiphenomenalism. Thus, epiphenomenalist theories are not difficult to find even in the modern wave of consciousness research. Certainly, they are much more common than interactionist theories. Yet, they are hardly any more convincing than interactionist dualism would be. The weakness of interactionism is that it cannot even begin to explain the mechanisms of mental causation, whereas the weakness of epiphenomenalism is that it flatly denies the existence of mental causation. As neither of these solutions is terribly convincing, perhaps we must look elsewhere for more credible theories.

It may be asked why epiphenomenalism should count as a variety of dualism at all. The principal reason is that all forms of epiphenomenalism postulate a mental realm that has no causal powers. Now, for something to be “real”, or to “exist” in the way that science could take it as a physically “real” phenomenon whose existence can in some way be detected or empirically tested, it needs to have at least *some* causal powers. Otherwise, it would be totally impossible to ever detect or confirm either the presence or the absence of the postulated entity if its existence makes no difference whatsoever in the happenings of the physical world. A phenomenon without any causal powers cannot be detected objectively or empirically, because there is no way it could influence any physical measurements or instruments, no matter how sophisticated. Therefore, any theory that postulates causally powerless entities, by definition of the empirically or physically real, thereby postulates nonphysical entities. This makes epiphenomenalism a version of dualism.

However, the nonphysical entities postulated by an epiphenomenalist theory may be viewed either as composed of a fundamentally nonphysical *substance* or as nonphysical *properties* of physical things. The former kind of theory represents *substance dualism* and the latter is often called *property dualism*. In fact, most of the modern versions of epiphenomenalism are closer to property dualism than substance dualism. They suggest that certain types of neural activities or computations – or perhaps quantum phenomena going on in the brain – have epiphenomenal features or properties. Our subjective psychological reality, thus, would consist of those nonphysical features emerging from the physical features and activities of the brain. Property dualism coupled with epiphenomenalism may come fairly close to strong emergent materialism, a theory that we will explore in more detail in Section 1.3.

## **Parallelism**

Causal interaction between the mental and the physical seems to be inexplicable in scientific terms. Epiphenomenalism gets rid of only one half of this problem by denying that the mental realm could causally influence the physical realm. But is it not equally mysterious how causal effects could go the other way around? How does neural activity in the brain get causal access to a nonphysical consciousness? The causal interface from the physical to the mental realm remains to be explained.