

The Functionalist Trilemma

Why Functional Equivalence Does Not Explain Consciousness

Wolfgang Stegemann

Abstract

Functionalism is among the most influential positions in the philosophy of mind. Its central idea is that mental states are determined not by their material constitution but by their functional or causal role. This gives rise to the possibility of multiple realization: if the same functional organization can be realized in biological and artificial systems, then, in principle, the same mental states should also be possible in different material substrates.

This paper argues that this inference does not hold for phenomenal consciousness. Functional equivalence initially establishes equivalence only with respect to functionally specified properties. For phenomenal equivalence to follow, an additional bridge premise is required according to which a particular functional organization is sufficient for experience. There are three argumentative possibilities for such a bridge. It may be stipulated definitionally, justified as an empirical generalization, or asserted as a fundamental psychophysical principle, for example as a necessary identity or a fundamental law. In the first case, consciousness is not explained but functionally redefined. In the second case, empirical evidence from biological systems is insufficient to establish substrate independence. In the third case, the transition from function to experience is carried by an additional principle rather than by the functional description itself.

This problem can be formalized as a factorization problem. A functional description is an abstraction that maps different realizations onto the same functional class. Functionalism about consciousness additionally claims that experience factors through precisely this abstraction, and therefore remains invariant within every functional equivalence class. This invariance does not follow from the abstraction itself. It is the central additional assumption of functionalism. Functionalism can therefore describe functional equivalence, but it cannot derive phenomenal equivalence from functional equivalence alone.

1. The Actual Problem of Functionalism

Functionalism rests on an extraordinarily compelling basic intuition. Whether a system calculates, navigates, plays chess, or answers linguistic questions does not seem to depend on the material from which it is made. The same function can be realized by very different material structures. A multiplication remains a multiplication whether it is performed by a human being, a mechanical calculator, or a computer.

This suggests the following thought: if mental states are likewise determined by their functional roles, then their material realization must also be secondary. A state would count as pain, for example, if it is caused by certain inputs, interacts with certain other states, and

produces certain behavioral consequences. What materially realizes that state would be irrelevant to its identity.

Compared with a simple identification of mental states with particular neural states, this position has an obvious advantage. If pain were identical with one highly specific neural state, a biologically different organism might be incapable of pain. Functionalism avoids this consequence. Different material states can realize the same functional role.

For beliefs, calculations, information processing, or certain forms of cognitive performance, this line of reasoning has considerable plausibility. Phenomenal consciousness, however, introduces an additional problem.

There is a difference between the claim

$$F(x)$$

and the claim

$$E(x)$$

where $F(x)$ means that system x realizes a particular function, and $E(x)$ means that there is something it is like for that system, that is, that the system undergoes experience.

The problem is not whether experience is connected with functions. In humans, it obviously is. The problem is what could make it follow from a functional description that the realization of this function is experienced.

At precisely this point, functionalism requires an additional premise.

2. Functional Equivalence

Consider two systems x and y . With respect to a specified function F , both realize the same organization. We can write:

$$x \equiv_F y$$

This equivalence can be extremely strong. It need not be limited to identical outward behavior. Internal state transitions, feedback loops, memory operations, self models, attentional systems, and response dispositions can all be included in the functional description.

From

$$x \equiv_F y$$

it initially follows only that x and y are equal with respect to the functional properties specified by F . It does not follow, for some other property P , that

$$P(x) = P(y)$$

A human being and a computer may perform the same addition. It does not follow that both are made of carbon, that both become hungry, age, or feel pain.

This is trivial. In the case of consciousness, however, precisely this trivial distinction becomes decisive. Let E denote the property of experience. Then functional equivalence alone does not establish:

$$x \equiv_F y \Rightarrow E(x) = E(y)$$

For this conclusion to be available, a further premise is required:

$$\forall x [F_C(x) \rightarrow E(x)]$$

where F_C denotes the functional organization that is supposed to be sufficient for consciousness.

Functionalism therefore cannot simply infer equivalence of consciousness from functional equivalence. It must explain why precisely this functional organization guarantees consciousness.

This is where the actual philosophical problem begins.

3. The Functionalist Trilemma

For the claim

$$\forall x [F_C(x) \rightarrow E(x)]$$

there are three argumentative possibilities. The connection between F_C and E can be fixed by definition. It can be justified as an empirical generalization from observed cases. Or it can be taken as a fundamental psychophysical principle, for example as a necessary identity or as a fundamental law connecting functional organization and experience.

This distinction concerns the status of the bridge. A fundamental psychophysical law may itself be discovered empirically. The crucial point, however, is that its universal validity would then not follow from the functional description. It would enter the theory as an additional basic principle. None of the three possibilities therefore yields the desired transition to experience from functional equivalence alone.

4. First Possibility: Consciousness Is Defined Functionally

Analytic functionalism may determine consciousness, or particular mental states, in terms of their functional roles. Then, for example:

$$E(x) := F_C(x)$$

Anything that realizes F_C satisfies the condition for E by definition.

The inference

$$F_C(x) \rightarrow E(x)$$

is now unassailable. It is true by definition.

The original problem, however, has not thereby been solved. The phenomenon to be explained was that in certain systems something is experienced at all. Pain is not merely a discriminable state that produces avoidance behavior, modifies memories, captures attention, and causes reports. Pain hurts.

If all functionally describable properties of pain are collected under F_C and F_C is then defined as pain, this tells us which functions we wish to group under the concept of pain. It does not show why their realization is experienced.

Definitional functionalism can therefore eliminate an explanandum by changing its designation.

Formally, the question

$$\text{Why } F_C(x) \wedge E(x)?$$

is transformed, by the definition

$$E(x) := F_C(x)$$

into the identity

$$\text{Why } F_C(x) \wedge F_C(x)?$$

The phenomenal difference has not been explained. It has been removed from the problem.

This argument is not directed against functional definitions as such. For many scientific purposes they are extremely useful. If consciousness is to be operationalized, observable or measurable criteria are required. The error arises only when successful operationalization is confused with an explanation of the phenomenon that has been operationalized.

A definition can specify when a predicate is to be applied. By itself, it cannot explain why the phenomenon designated by that predicate exists.

The Eliminativist Response

A radical functionalist may explicitly accept this consequence. Such a view may deny that, beyond functionally describable discriminations, dispositions, and self attributions, there remains an independent phenomenal explanandum. Dennett's critique of traditional notions of qualia is a prominent version of this strategy (Dennett 1991). A conceptual qualification is important here. One need not deny that people have pains or experience colors. What is denied is that there are, in addition, intrinsic qualia independent of all functional and representational roles.

Such a position does not refute the trilemma. It chooses its first horn. The phenomenal explanandum is wholly transferred into a functional description. The functionalist can then consistently explain why a system reports pain, avoids it, remembers it, or discriminates it.

But the view does not additionally answer why these processes are experienced. Instead, it denies that any independent question remains once the functional explanation is complete.

The controversy has therefore shifted. Anyone who accepts phenomenal experience as an independent explanandum does not obtain an explanation of that explanandum from a functional redefinition. Anyone who does not accept it adopts an eliminativist or illusionist position. This is philosophically possible, but it does not rescue functionalism as an independent explanation of phenomenal experience. It resolves the problem by denying that the problem exists in the assumed form.

5. Second Possibility: The Bridge Is Empirical

The functionalist may avoid the definitional strategy. The view may concede that experience is not identical with its functional description while claiming that certain forms of functional organization empirically and reliably give rise to consciousness.

We then obtain

$$F_C(x) \rightarrow E(x)$$

as an empirical regularity.

The immediate question concerns the evidential basis. In humans, we observe a connection among certain biological, neural, and functional states and conscious experience. In other animals, too, warranted attributions can be made on the basis of biological homologies, behavior, neurophysiological structures, and evolutionary continuities.

The empirical starting point therefore has approximately the following structure:

$$F_C(O) \wedge B(O) \wedge E(O)$$

where O denotes a particular organism and B its biological organization.

Functionalism, however, seeks to derive:

$$\forall x [F_C(x) \rightarrow E(x)]$$

The biological condition is thereby dropped. This step is not justified by the initial data.

From

$$F_C(O) \wedge B(O) \wedge E(O)$$

we cannot immediately determine, either logically or empirically, whether F_C alone is sufficient or whether

$$F_C \wedge B$$

is required for the occurrence of E .

The data underdetermine this alternative. As long as consciousness is demonstrated, or justifiably attributed, only in systems in which functional organization and a particular biological organization occur together, the evidence does not establish that biological organization is irrelevant.

The problem has the structure of an ordinary confound. We observe:

$$A \wedge B \rightarrow C$$

and subsequently claim:

$$A \rightarrow C$$

without determining the independent contribution of B .

Precisely because functionalism wishes to treat the material substrate as irrelevant, it must show that the substrate is irrelevant. It cannot simply remove it from the description before that question has been examined.

6. Multiple Realization Does Not Establish Multiple Realization of Experience

At this point, the thesis of multiple realizability becomes decisive.

Many functions are demonstrably multiply realizable. Addition can be realized neurally, electronically, or mechanically. A control circuit can consist of biological cells, electronic components, or hydraulic elements. A particular input output transformation can occur in very different physical substrates.

This establishes:

$$\text{MR}(F)$$

that is, multiple realizability of a function.

Functionalism about consciousness, however, requires:

$$\text{MR}(E)$$

that is, multiple realizability of experience.

From

$$\text{MR}(F)$$

it does not follow that

$$\text{MR}(E)$$

unless one already assumes:

$$F_c \rightarrow E$$

The argument thereby becomes circular. The multiple realizability of functional organization is supposed to show that consciousness is substrate independent. Yet the substrate independence of consciousness is needed in order to infer the multiple realizability of consciousness from the multiple realizability of functional organization.

Formally:

$$\text{MR}(F_C)$$

and

$$F_C \rightarrow E$$

may yield

$$\text{MR}(E)$$

But the crucial second premise is precisely what must be justified.

Multiple realizability of functions therefore cannot establish multiple realizability of phenomenal consciousness.

7. Third Possibility: The Bridge Is an Identity or a Fundamental Psychophysical Law

The strongest version of functionalism can go further. It need not claim that functional organization and consciousness are merely empirically correlated. It may claim that a sufficiently precise functional organization necessarily determines experience.

Then:

$$\text{Nec}(F_C(x) \rightarrow E(x))$$

or, more strongly:

$$F_C = E$$

The advantage is obvious. If such a necessity obtains, the substrate indeed plays no decisive role. But the question then becomes: how do we know that this necessity obtains?

It cannot follow from functional equivalence itself. Otherwise, the conclusion would already be presupposed. Nor is the intuition that a functionally identical system must obviously be conscious sufficient, since that intuition already assumes the functionalist invariance whose validity is under examination.

An obvious objection appeals to a posteriori necessity. Since Kripke, it has been familiar that an identity may be necessary even though it is known only empirically. The classic example is:

$$\text{Water} = H_2O$$

The functionalist might accordingly claim that

$$F_C = E$$

is likewise a necessary identity that need not be derivable a priori but may instead be discovered empirically (Kripke 1980). This possibility cannot be excluded by logic alone. It helps functionalism, however, only if it can be shown empirically that F_C and E in fact designate the same property. That demonstration is precisely what is missing. Existing evidence does not isolate functional organization from biological realization. It shows F_C , biological organization, and experience occurring together.

The analogy with water and H_2O therefore does not remove the burden of justification. An a posteriori identity is not established merely because two descriptions regularly co-occur or because one has a similar causal role. In the case of consciousness, it would have to be shown that the same functional organization, independently of the realizing substrate, is identical with the same experience.

Thus:

$$\text{a posteriori possible} \neq \text{a posteriori established}$$

A second version dispenses with identity altogether. A nonreductive position may posit a fundamental psychophysical law connecting functional organization and experience. Let such a law be denoted by L_{FE} . We then have, for example:

$$L_{FE}: F_C \mapsto E$$

This possibility does not break the trilemma either. Instead, it makes explicit what is additionally required. F_C alone still does not entail E . The transition emerges only together with L_{FE} :

$$F_C + L_{FE} \Rightarrow E$$

Such a law may be treated as fundamental and perhaps empirically discovered. The explanatory burden, however, then lies with L_{FE} . Functional organization describes one side of the relation, while the psychophysical law supplies the bridge to experience. Functionalism by itself has not generated this bridge.

The third possibility is therefore not a counterexample to the trilemma but its strongest form. Either functional organization is identified with experience, or an additional fundamental law connects the two. In either case, the decisive transition does not come from functional equivalence itself.

8. Organizational Invariance and Dancing Qualia

A particularly strong version of the functionalist idea appears in Chalmers's thesis of organizational invariance. According to this thesis, a system with the same sufficiently fine grained functional organization should also possess the same kind of conscious experience. The well known Dancing Qualia argument seeks to support this thesis by means of a

gradual replacement of biological components with functionally equivalent components (Chalmers 1996).

Suppose that experience changes during this replacement even though functional organization is preserved completely. Color experiences, for example, might shift while all reports, judgments, memories, and reactions remain unchanged. The system would continue to say that everything looks exactly as it did before. For Chalmers, this produces substantial conceptual tension. If experience changes dramatically, the system should be able to notice that change. If every functional reaction remains identical, the system instead appears to be fundamentally mistaken about its own phenomenal state.

This argument is stronger than the mere observation that Dancing Qualia are counterintuitive. It nevertheless contains a bridge premise of its own. For a phenomenal change to generate incoherence, it must be the case that a relevant change in experience necessarily produces a functionally accessible change in noticing, judging, or reporting.

Let J denote the functionally accessible cognitive response to one's own phenomenal state. The argument then requires a principle of the form:

$$\Delta E \rightarrow \Delta J$$

The functional invariance stipulated in the thought experiment, by contrast, yields:

$$\Delta J = 0$$

Only given the first bridge rule does it follow that:

$$\Delta E = 0$$

But this bridge rule is not contained in the mere preservation of functional organization. It already says that phenomenally relevant differences must necessarily be functionally accessible. If noticing is understood entirely in functional terms, the assumption becomes: every phenomenal difference relevant to consciousness must be reflected in a functional difference within the system.

This may be true. But it is itself a psychophysical coherence assumption connecting experience and functional accessibility. Anyone who accepts it can rule out Dancing Qualia. Organizational invariance has then not been derived from functional organization alone, but from functional invariance plus an additional rule concerning the relation between experience and cognition.

The argument therefore yields a conditional result:

$$(\Delta E \rightarrow \Delta J) \wedge (\Delta J = 0) \Rightarrow \Delta E = 0$$

The logical form is valid. What remains contested is the first premise. It is located precisely where the trilemma identifies the missing bridge.

This is especially revealing for a nonreductive position. If functional organization and experience are not identical, additional psychophysical principles must explain why particular functional states are connected with particular experiences. Dancing Qualia may increase the plausibility of such a coherence principle. It does not replace that principle.

Organizational invariance therefore confirms the central structure of the problem:

$$F_C$$

by itself does not yet yield

$$E$$

The connection must additionally be justified, discovered, or posited as a fundamental principle.

9. The Abstraction Problem

The problem can be stated more generally. A functional description is necessarily an abstraction.

Let S be the set of possible concrete systems and let

$$A: S \rightarrow F$$

be a functional mapping. A assigns to each concrete system its functional description.

Two systems x and y count as functionally equivalent if

$$A(x) = A(y)$$

This defines an equivalence relation:

$$x \sim_F y \Leftrightarrow A(x) = A(y)$$

The functional description therefore generates classes of systems that are not distinguished with respect to the selected functions.

This is precisely its strength. A biological nervous system and a silicon machine can belong to the same functional class even though they differ substantially in their physical constitution.

In the interesting case, the mapping A is therefore noninjective:

$$A(x) = A(y)$$

although

$$x \neq y$$

The functional description contains less information than the concrete system.

This is precisely what gives rise to the decisive problem.

10. Consciousness Would Have to Factor Through Functional Abstraction

Now let

$$E: S \rightarrow P$$

assign phenomenal properties to systems.

For functional organization to determine consciousness completely, the following would have to hold:

$$A(x) = A(y) \Rightarrow E(x) = E(y)$$

Experience would therefore have to remain invariant within every functional equivalence class generated by A .

Mathematically, E would have to factor through A . There would have to exist a mapping

$$g: F \rightarrow P$$

such that

$$E = g \circ A$$

The following two statements are equivalent:

$$\exists g: F \rightarrow P \text{ such that } E = g \circ A \Leftrightarrow \forall x, y \in S [A(x) = A(y) \Rightarrow E(x) = E(y)]$$

This is the most precise formulation of the functionalist claim. Consciousness then depends exclusively on functional class. Everything lost by A in the process of abstraction is irrelevant to E .

But this factorization does not follow from A .

From

$$A(x) = A(y)$$

we can conclude only that x and y are equal with respect to the properties captured by A . Whether

$$E(x) = E(y)$$

holds depends on whether E is invariant under all differences removed by A .

That is an additional claim about E .

Functional abstraction cannot derive that claim from itself, because it does not contain precisely those differences whose possible relevance would need to be examined.

This is the fundamental epistemic problem of functionalism.

11. Substrate Irrelevance Is Produced by Omitting the Substrate

The same process can be described more intuitively. A biological organism O possesses a plurality of properties:

$$O = \{F, B, H, D, \dots\}$$

Here, F stands for functional organization, B for biological organization, H for developmental history, and D for concrete dynamics.

Functional abstraction maps this system to:

$$A(O) = F$$

Now a machine M is constructed for which likewise:

$$A(M) = F$$

It follows that:

$$A(O) = A(M)$$

and therefore:

$$O \sim_F M$$

This is entirely correct.

The functionalist then wishes to claim that differences in B , H , D , and further realization properties are irrelevant to consciousness. But this conclusion cannot be obtained from $A(O) = A(M)$. The mapping A was constructed precisely so that these differences are not taken into account.

The central error can therefore be summarized in one sentence:

Functionalism does not establish the irrelevance of the substrate. It first produces that irrelevance through abstraction and then treats the result of the abstraction as a property of the object itself.

This confuses epistemic abstraction with ontological irrelevance.

From

$$A(x) = A(y)$$

it follows only that:

x and y are indistinguishable under A .

It does not follow that:

x and y differ in no property relevant to E .

The second statement is the functionalist thesis itself. It therefore cannot be justified by the first.

12. The Granularity Problem

A functionalist might reply that the functional description is simply not detailed enough. All causally relevant states of the system must be included. With a sufficiently fine grained functional description, the apparently missing properties would themselves be represented functionally.

This response can be taken further. A modern conception of function need not be restricted to the input output relations of a computer. It can include embodiment, recurrent dynamics, metabolism, self regulation, development, or even autopoietic self production in the functional description. An expanded specification might therefore take the form:

$$F_C^* = \{F, \text{body, dynamics, metabolism, development, self production, ...} \}$$

This does not defeat the abstraction argument. It merely shifts the boundary of abstraction. If biological and historical properties are themselves redescribed functionally, the question remains which concrete properties are lost in this functionalization and whether precisely those properties are constitutive of experience. If, by contrast, they are specified so concretely that their biological realization becomes part of the condition itself, the space of possible realizations contracts.

Autopoiesis makes this alternative especially clear. If autopoiesis is described merely as the abstract role of self maintenance, it may perhaps be realized by very different systems. But then it has not been shown that this abstract role preserves all properties of biological self production that might be relevant to consciousness. If autopoiesis is instead specified as concrete chemical and organismic self production, the description is no longer substrate independent in the same strong sense.

This generates a further dilemma.

If the functional description remains coarse enough to place very different material systems in the same functional class, realization differences necessarily remain outside the description. It therefore remains open whether any of those differences is constitutive of consciousness.

If the functional description is progressively refined until all potentially relevant dynamic, bodily, biological, and historical properties are included, it approaches the concrete realization itself. The thesis of strong substrate independence then increasingly loses its content.

Functionalism therefore faces a granularity alternative. An abstract functional description permits multiple realization, but it cannot establish the irrelevance of the properties from which it abstracts. A maximally concrete description might capture relevant properties, but it loses precisely the abstraction from which the thesis of multiple realizability derives its force.

Functionalism cannot guarantee both maximal realization independence and maximal explanatory completeness through the same abstraction. The more concrete organization it includes in F_C , the less remains of the original claim that an abstract functional role alone, independently of its realization, is sufficient for consciousness.

13. Why Simulation Is Not Enough

The problem becomes especially clear in artificial systems. A machine can realize a functional structure abstracted from a biological system. It can calculate, learn, plan, report, model itself, allocate attention, and respond to internal states.

Each of these capacities establishes a functional property. Their combination likewise remains, initially, a functional property.

Suppose a machine satisfies all functional tests that we have specified for consciousness:

$$F_1(M) \wedge F_2(M) \wedge \dots \wedge F_n(M)$$

Then:

$$F_c(M)$$

If the tests were defined functionally, this conclusion is even necessary.

The question

$$E(M)?$$

remains logically distinct.

It can be answered only if it has additionally been shown that:

$$F_c(x) \rightarrow E(x)$$

A perfect functional test for consciousness would therefore also be a perfect test for experience only if functionalism were already true. The test cannot independently confirm the theory if its validity presupposes the theory.

14. Behavior Does Not Solve the Problem

The same applies to behavior. The more similarly a machine behaves to a human being, the stronger the intuitive impression that the two should also be similar inwardly.

This intuition is understandable. In everyday life, we constantly infer the experience of other human beings from their behavior. But there the inference has a background that is precisely what is missing in the machine case.

Other human beings belong to the same biological species. They possess comparable nervous systems, comparable bodies, comparable development, similar physiological responses, and a shared evolutionary origin.

The inference from my own experience to the experience of another human being therefore does not amount merely to:

$$\text{same behavior} \rightarrow \text{same experience}$$

It rests on a much more extensive structural similarity.

In the machine case, these commonalities are removed while functional and behavioral similarities are preserved. Under those conditions, one can no longer appeal to the same inference.

The greater the material and genetic difference between two systems, the stronger the justification required for the claim that only the preserved functional structure is relevant.

15. Function Is a Relational Description

The concept of function itself sharpens the problem. A function designates a role within a context. The heart pumps blood. A neuron transmits signals. A computer computes a function. A particular state leads, under specified conditions, to other states.

A functional description therefore answers questions of the form:

What does x do?

or:

What role does x have within the system?

Consciousness, however, initially raises a different question:

What state is the system as a whole in?

If consciousness is a state of a particular organization rather than an additional function, it may have functional consequences without being exhaustively determined by a description of those consequences.

A fever state likewise has numerous functions and consequences. A functional description of those consequences is not identical with a physiological explanation of fever.

Similarly, experience may affect behavior, reports, memory formation, attention, and decisions without being identical with the set of these functional relations.

Functionalism reverses this direction. It takes the relational consequences and uses them to define the identity of the state. For some states this may be appropriate. In the case of consciousness, this is precisely what has to be demonstrated.

16. From Description to Ontology

The error can now be formulated more generally. Functionalism begins as a method of description. It maps systems according to particular relations and ignores differences that are irrelevant to the function in question. This is scientifically entirely legitimate.

The problematic step comes afterward:

epistemically unrepresented

becomes

ontologically irrelevant

This inference is invalid.

A map contains neither soil chemistry nor smells, microorganisms, or the history of the landscape. The fact that two landscapes look identical on a particular map does not establish that their unrepresented properties are irrelevant to every possible question.

The same applies to a functional map of an organism. It may be complete for a particular task. It does not follow that it is complete for explaining every property of the organism.

The decisive question is therefore not: Is the functional description complete with respect to the function? It may well be.

The decisive question is: Is the functional description complete with respect to the conditions under which experience arises?

This second claim requires independent justification.

17. Genesis as an Excluded Variable

A biological organism is not an arbitrary carrier of an abstract function. It is the present result of a historical process of self production, development, and evolutionary selection.

Its current organization did not arise independently of this process. Nervous system, metabolism, bodily regulation, plasticity, and behavior developed within a self maintaining system.

A functional snapshot necessarily abstracts from parts of this genesis.

For many functions this is unproblematic. To perform an addition, a machine need not have the same history as a human being.

But it does not follow that consciousness can likewise be reproduced independently of its genesis.

This question remains empirically and theoretically open as long as the constitutive conditions of consciousness have not been determined.

If experience were, for example, bound to the organization of a self producing living system, an artificial machine might reproduce every relevant observable function without reproducing the constitutive condition of the state.

Functional equivalence cannot exclude this possibility. Genesis has already been eliminated in the functional abstraction.

18. Construction Is Not a Reconstruction of Genesis

Here lies a further difference between biological and artificial systems. A machine can be constructed according to a functional specification. An organism, by contrast, emerges through a historical process whose present state is a condensation of its development.

A functional map can be obtained from a biological system:

$$O \rightarrow F$$

A different system can then be constructed:

$$F \rightarrow M$$

But it does not follow that:

$$O \rightarrow F \rightarrow M = O$$

The second mapping does not invert the first.

If the abstraction

$$O \rightarrow F$$

has lost information, that information cannot be recovered from F .

Formally, the functional mapping is typically noninjective:

$$A(O_1) = A(O_2)$$

while

$$O_1 \neq O_2$$

It therefore has no unique inverse.

A machine constructed on the basis of the map reproduces the map. It does not necessarily reproduce the landscape from which the map was derived.

If consciousness belongs to those properties of the landscape that were not fully preserved when the map was produced, its existence cannot be inferred from agreement between the maps.

19. The Fundamental Functionalist Problem as a Proposition

The entire objection can now be summarized in a simple structure.

Let

$$A: S \rightarrow F$$

be a functional abstraction of concrete systems, and let

$$E: S \rightarrow P$$

assign phenomenal states.

Functionalism about consciousness requires:

$$A(x) = A(y) \Rightarrow E(x) = E(y)$$

for all relevant x, y .

Exactly then there exists a mapping

$$g: F \rightarrow P$$

with

$$E = g \circ A$$

if E is constant on every fiber of A . Formally:

$$\exists g: F \rightarrow P \text{ such that } E = g \circ A \Leftrightarrow \forall x, y \in S [A(x) = A(y) \Rightarrow E(x) = E(y)]$$

The existence of A , however, does not prove the existence of g .

Nor does the fact that functional and phenomenal properties co-vary within a subset of biological systems establish this factorization across the entire set of possible realizers.

It follows that:

Functional equivalence does not establish phenomenal invariance.

Phenomenal invariance must be shown, posited, or defined additionally.

If it is defined, the explanandum is functionalized. If it is asserted empirically, independent evidence for substrate independence is lacking. If it is posited metaphysically or nomologically, that additional principle carries the explanatory burden rather than functionalism itself.

20. What Remains of Functionalism

Nothing in this argument implies that functional analysis is worthless. The opposite is true.

Functional description is indispensable for analyzing behavior, cognition, information processing, control, or interaction. It can explain how a system discriminates, categorizes, stores, computes, reports, and acts.

The error consists only in inferring ontological completeness from this explanatory strength.

For functional properties, it may hold that:

$$A(x) = A(y) \Rightarrow F(x) = F(y)$$

because A was constructed precisely for those properties.

For a property whose functional completeness is itself in dispute, the same inference cannot be presupposed.

Functionalism may therefore be an excellent theory of functional cognition. As a theory of phenomenal consciousness, it is incomplete.

It tells us what organization is supposed to realize consciousness. From that organization alone, however, it cannot justify why anything is experienced at all.

21. Conclusion

The appeal of functionalism rests on a correct insight. Functions can be realized by very different material systems. This accounts for the enormous technical and theoretical power of functional abstraction.

The error begins only when this insight is transferred to consciousness.

That two systems have the same functions establishes their equality with respect to those functions. It does not establish their equality with respect to every property.

Consciousness therefore requires a bridge premise:

$$F_C(x) \rightarrow E(x)$$

This premise cannot be justified by functional equivalence itself.

If it is stipulated definitionally, it does not explain experience. If it is asserted empirically, the irrelevance of the substrate must be established empirically. If it is posited as a necessary identity or a fundamental psychophysical law, the decisive explanatory work lies in that additional principle.

The deeper error becomes visible through functional abstraction. A concrete system is mapped onto a functional class. Differences among realizations are deliberately removed. Equality within that abstraction is then taken to imply that the removed differences are irrelevant to consciousness.

An epistemic operation is thereby transformed into an ontological claim.

Functionalism does not establish the irrelevance of the substrate. It presupposes it.

Its central thesis about consciousness therefore cannot be derived from functional formalism. It has to be added to it.

This marks a clear boundary:

Function can explain function.

It does not follow that function is experienced.
--

As long as functionalism provides no independent justification for this second transition, it is not a sufficient explanation of consciousness. It is a description of those functions whose realization is already assumed to be capable of being conscious.

References

Block, Ned. 1978. "Troubles with Functionalism." In C. Wade Savage, ed., *Perception and Cognition: Issues in the Foundations of Psychology*, Minnesota Studies in the Philosophy of Science, vol. 9. University of Minnesota Press.

Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Dennett, Daniel C. 1991. *Consciousness Explained*. Little, Brown and Company.

Kripke, Saul A. 1980. *Naming and Necessity*. Harvard University Press.

Putnam, Hilary. 1967. "Psychological Predicates." In William H. Capitan and Daniel D. Merrill, eds., *Art, Mind, and Religion*, pp. 37–48. University of Pittsburgh Press.