How Constant are Fundamental Physical Quantities?

W. Eichendorf and M. Reinhardt
Astronomisches Institut der Universität Bochum

(Z. Naturforsch. 32 a, 532—537 [1977]; received April 1, 1977)

We reinvestigate Dirac’s large number hypothesis (LNH) which implies the variation of one or more basic physical constants with time. We show that the ratio of the inertial masses of elementary particles and the fine structure constant \( \alpha \) do not vary with time in the LNH.

I. Dirac’s Large Number Hypothesis (LNH)

In general it is a tacit assumption in physics and particularly in cosmology that physical laws are the same in every point of space-time, i.e., that they do not change with position in space or the age of the universe. This implies that everywhere and at all times their form and the numerical content are the same.

So far the former has never been questioned and there are good reasons for not doing this. First we are able to interpret the electromagnetic spectra of galaxies up to redshifts \( z \approx 0.5 \) and of quasars up to \( z \approx 4 \). This means that at least the laws of atomic physics found in terrestrial laboratories are valid over 80 to 90 percent of the age of the universe and of the radius of the observable universe. Even if one doubts that the redshifts of quasars are cosmological, as some people do, we are on firm ground using only galaxies, and the number given above changes to 35 to 50 percent. Secondly, if one would drop the assumption of the invariability of the form of physical laws, speculation would get completely out of hand.

However, the hypothesis that fundamental physical constants are a function of cosmic time has been put forward repeatedly. In this paper we restrict our attention mainly to the most popular of these proposals, namely Dirac’s large number hypothesis (LNH). Dirac \(^1, 2\) expressed the opinion that large dimensionless numbers like the ratio of the electromagnetic force and the gravitational attraction between protons and electrons

\[
\frac{e^2}{G m_p} = 2.27 \cdot 10^{39} \\
\frac{e^2}{G m_e^2} = 4.17 \cdot 10^{42} \\
\frac{e^2}{G m_p^2} = 1.24 \cdot 10^{56}
\]

(\( e \) — elementary charge, \( G \) — gravitational constant, \( m_p \) and \( m_e \) — the masses of a proton and an electron, respectively) reflect in some way the state of the universe. He based this quasi-Machian argument (see e.g., Reinhardt \(^3\)) on the approximate coincidence of the numerical value of (1) with the ratio of the age of the universe \( t_0 \) and the units of time that can be constructed from atomic constants (Planck’s constant)

\[
t_0 / (e^2/m_e c^3) = (1.06 \cdot 10^{23} \text{ s}^{-1}) t_0,
\]

\[
t_0 / (e^2/m_p c^3) = (1.95 \cdot 10^{38} \text{ s}^{-1}) t_0,
\]

\[
t_0 / (h/m_e c^2) = (1.24 \cdot 10^{30} \text{ s}^{-1}) t_0,
\]

\[
t_0 / (h/m_p c^2) = (2.27 \cdot 10^{28} \text{ s}^{-1}) t_0.
\]

The age of the universe is of the order of \( H_0^{-1} \), where \( H_0 \) is the Hubble constant, for most cosmological models. Exceptions are the steady state theory, which can be ruled out almost certainly because of the blackbody nature of the microwave background radiation and the evolution of extragalactic radio sources with cosmic time, the Lemaître models and a few exotic models. For Lemaître models \( H_0^{-1} \) is still a fair estimate of the time in which the mean energy density of the universe changes markedly. Dirac assumes that the distance of galaxies is proportional to time \( t^4 \). Then the age
of the universe is exactly the inverse of the Hubble constant, as in the empty Friedmann universe.

Estimates of the present expansion rate of the universe range from $H_0 = 50$ to $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Using the lower value $H_0 = 6.36 \times 10^{17} \text{ cm s}^{-1} \text{ Mpc}^{-1}$ and (2) to (5) become $6.74 \times 10^{17}$, $1.24 \times 10^{17}$, $7.86 \times 10^{17}$, and $1.44 \times 10^{17}$, respectively. In 1937, when Dirac put forward his hypothesis, $H_0$ was thought to be $530 \text{ km s}^{-1} \text{ Mpc}^{-1}$, so that (2) to (5) would have been smaller by a factor 0.094. Dirac did not demand an exact numerical coincidence between any of these large dimensionless numbers, only a rough coincidence of the order of magnitude, being even here rather generous. He proposed that "any two of the very large dimensionless numbers occurring in Nature are connected by a simple mathematical relation, in which the coefficients are of order unity". This statement is known as Dirac's LNH. In its simplest form, the only ever seriously discussed, it implies that the ratios of all large dimensionless numbers are time-independent. If one follows this hypothesis, one or more of the fundamental physical quantities occurring in (1) to (5) must change with time in order to check the increase of the age of the universe and to keep the ratio of (1) and each of (2) to (5) constant.

Combining the three ratios (1) with (2) to (5), taking the time derivative and dividing by $t_0$ we obtain:

$$t_0^{-1} = 4 \dot{e}/e - G/G - 2 \dot{m}_e/m_e - \dot{m}_p/m_p - 3 \dot{c}/c,$$  (6.1)

$$= 4 \dot{e}/e - G/G - \dot{m}_e/m_e - 2 \dot{m}_p/m_p - 3 \dot{c}/c,$$  (6.2)

$$= 2 \dot{e}/e - G/G - \dot{m}_e/m_e - 2 \dot{m}_p/m_p - 2 \dot{c}/c + \dot{h}/h,$$  (6.3)

$$= 2 \dot{e}/e - G/G - \dot{m}_e/m_e - 2 \dot{c}/c + \dot{h}/h,$$  (6.4)

$$= 4 \dot{e}/e - G/G - 3 \dot{m}_e/m_e - 3 \dot{c}/c,$$  (6.5)

$$= 4 \dot{e}/e - G/G - 3 \dot{m}_e/m_e - 3 \dot{c}/c,$$  (6.6)

$$= 2 \dot{e}/e - G/G - 3 \dot{m}_e/m_e - 2 \dot{c}/c + \dot{h}/h,$$  (6.7)

$$= 2 \dot{e}/e - G/G - 3 \dot{m}_e/m_e - 2 \dot{c}/c + \dot{h}/h.$$  (6.8)

From (6.1) and (6.2) follows

$$\dot{m}_p/m_p - \dot{m}_e/m_e = 0.$$  (7)

Equation (7) reduces (6.2) to (6.5) and (6.4) to (6.7). Then combining (6.5) and (6.7) results in

$$2 \dot{e}/e = \dot{h}/h - \dot{c}/c = 0,$$  (8)

$$2 \dot{h}/h - G/G - 3 \dot{m}_e/m_e - \dot{c}/c = t_0^{-1},$$  (9)

$$2 \dot{h}/h - G/G - 3 \dot{m}_e/m_e = t_0^{-1}.$$  (10)

Thus Dirac's hypothesis boils down to Eqs. (7), (8), (9), and (10) to be valid simultaneously. Note that only two quantities, namely the gravitational coupling constant and the mass of elementary particles could vary alone with time in Dirac's hypothesis. In particular (7) and (8) imply that the ratio of electron mass to proton mass $m_e/m_p$ and the fine structure constant $a = e^2/\hbar c$ should not vary with time.

II. Upper Limits on the Variation of $G$

Dirac proposed that the gravitational constant decreases in proportion to the age of the universe

$$-G/G = t_0^{-1}$$  (11)

and that all other quantities occurring in (7) to (10) are constant. A variation of $G$ would affect not only cosmology (e.g. general relativity would be violated) but also the solar system. Thus all bodies in the solar system held together by gravitational forces would expand, the radii of the planetary orbits $R$ would increase as $t^{-1}$ due to conservation of angular momentum, and the luminosity $L$ of the sun would have been higher in the past ($L \sim t^{-7}$).

The best radar measurements of the orbit of Mercury give an upper limit to $G/G$ of $4 \cdot 10^{-10} \text{ a}^{-1}$. This is not enough to rule out (11) with $t_0^{-1} \approx 5 \cdot 10^{-11} \text{ a}^{-1}$. An argument against a substantial variation of $G$ with time based on stellar evolution was brought up by Pochoda and Schwarzschild and independently by Gamow. They could show that the sun would have burned all its hydrogen in the core after roughly $1.5 \cdot 10^9$ a of its life, if (11) were valid, the initial hydrogen content would have been $X = 0.76$ by mass, and if the age of the universe were $t_0 = 1.5 \cdot 10^{10}$ a. After this time the sun would have become a red giant which is clearly at variance with observations. The duration of the hydrogen burning phase is extremely sensitive to $X$. Thus with (11), the same $t_0$ and $X = 0.81$ the sun
would not have left the main sequence, if it were $4.5 \cdot 10^9 \text{a}$ old. Since the initial hydrogen content of the sun must have been $X \leq 0.72$ due to He-production in the early universe (see e.g. 9) and the age of the sun should be greater or equal to the age of the earth ($4.5 \text{ to } 6 \cdot 10^9 \text{a}$), a variation of $G$ according to (11) can be excluded, notwithstanding a possible increase of $t_0$ to $2 \cdot 10^{10} \text{a}$. Of course a new calculation of the evolution of the sun with a varying gravitational coupling constant would be desirable. At present it is difficult to give a firm upper limit on $G/G$ from this argument, albeit a variation according to Dirac’s hypothesis (11) seems to be excluded.

An upper limit on a variation of $G$ can be derived from an argument originally due to Teller 6. If (11) holds, the surface temperature of the earth would have been higher in the past due to the higher luminosity of the sun ($L \sim G^7$) and the smaller radius of the orbit of the earth ($R \sim G^{-1}$). Assuming a constant albedo of the earth the surface temperature would vary as

$$T \sim (L/R^2)^{1/4} = t^{-9/4}. \quad (12)$$

Teller’s argument proceeds then as follows. With a surface temperature of $288 \text{ K}$ at present and an age of the universe of $1.9 \cdot 10^9 \text{a}$ (the value fashionable in 1948), the surface temperature would have reached the boiling point of water $2.1 \cdot 10^9 \text{a}$ ago. This is clearly at variance with paleontological evidence. At this time we had a highly evolved life, including reptiles and fur trees, which could not survive temperatures around 373 K.

However, since then the estimates of $H_0$ have decreased and hence upper limits to the age of the universe increased by a factor of more than 10, thus making life not uncomfortable even for trilobites in the cambrian at a temperature of 309 K, if Dirac were right. Nevertheless Teller’s argument does not break down. We know nowadays that in the Swartkoppie-Hornstein strata of the Onverwacht formation in South Africa, $\Delta t = 3.4 \cdot 10^9 \text{a}$ ago in the precambrian, there existed some primitive life forms at a temperature around 333 to 363 K 12. Also we have temperature estimates based on the isotope abundances of hydrogen and oxygen in flint stone and cherts of strata of this epoch in North America, England and Africa. They give similar upper limits to the surface temperature of the earth $3.4 \cdot 10^9 \text{a}$ ago of 333 to 368 K 12. On the other hand (12) would imply with $t_0 = 2 \cdot 10^{10} \text{a}$ that the surface temperature would have reached 373 K only $2.2 \cdot 10^9 \text{a}$ ago. This clearly contradicts the observations 13.

Using the geochemical data mentioned above we can put a firm upper limit on a possible variation of $G$. Let us suppose that

$$G \sim t^n \quad (13)$$

then the surface temperature of the earth would vary like

$$T \sim t^{n/14} \quad (14)$$

which gives

$$n \leq 4/9 \log (T/T_0)/\log [(t_0 - t)/t_0]. \quad (15)$$

With $T < 333$ to 368 K and $\Delta t = 3.4 \cdot 10^9 \text{a}$ we get

$$-n < 0.34 \text{ to } 0.59 \quad (16)$$

and

$$-G/G < (1.7 \text{ to } 2.9) \cdot 10^{-11} \text{a}^{-1}. \quad (17)$$

An $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ leads to values still about 7 percent lower. Thus Dirac’s conjecture (11) is excluded by these limits.

Our limit is derived under the assumption that the albedo of the earth $3.4 \cdot 10^9 \text{a}$ ago was the same as today. In order to save Dirac’s hypothesis one has to assume a “coldhouse” effect of the archean atmosphere, i.e. the archean atmosphere should have reflected so much of the sun’s radiation, that the surface temperature of the earth would have been lower than that derived from the thermodynamical argument above. Detailed calculations of the radiative balance in a hypothetical archean atmosphere, probably mainly composed of water, methane, ammonia, and carbon dioxide, for a solar spectrum and luminosity of the sun as it is now according to Dirac’s hypothesis would be necessary in order to investigate this point. This is clearly beyond the scope of this paper. However, it seems that one would rather expect a “greenhouse” effect with the kind of archean atmosphere mentioned above 14. This would make the case for Dirac’s hypothesis even worse.

Recently two other lower limits on the variation of $G$ have been published. Mansfield 15 derived 1976 an upper limit

$$G/G \leq - (5.8 \pm 1) \cdot 10^{-11} \text{a}^{-1} \quad (18)$$

from the slowing down of the periods of pulsars using a polytrope model for neutron stars. Blake 16 (1977) using fossil evidence for possible variations of the duration of the solar day and the synodic
month over the last $4 \cdot 10^8$ a from the growth rhythms of bihalves and corals claims

$$|G/G| \lesssim 2 \cdot 10^{-11} \text{ a}^{-1}.$$  \hfill (19)

Blake’s limit is of the same order as ours, however it is fraught with the uncertainty of the change of the earth’s moment of inertia with changing $G$. Also the fossil data (counting of growth lines of corals) and their evaluation are more complicated and therefore more open to errors than the simple temperature estimate we use. Therefore we think that our limit (17) is still the best evidence against a variation of $G \sim t^{-1}$ alone.

Recently Dirac modified his LNH by assuming in addition continuous creation of matter. He argues that the total amount of matter in the universe $M$ expressed in units of the proton rest mass is a very large dimensionless number of the order of $10^{80}$. This is roughly the square of the age of the universe expressed in atomic units. He claims that therefore the total mass of the universe $M$ should vary as $t^2$. This would require continuous creation. In fact with $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and estimates of the mean matter density at the present epoch of $\rho_0 = 4.4 \cdot 10^{-30}$ to $4.4 \cdot 10^{-32} \text{ g cm}^{-3}$, the number of protons and neutrons in an euclidean sphere with radius $R = c t_0 \approx c H_0^{-1}$ is of the order of $10^{78}$ to $10^{80}$. Note that the concept of the total amount of matter in the universe is ill defined. E.g. if the universe is open, the total amount of matter in the universe is infinite. The numbers quoted above refer only to the observable universe. However that might suffice for a quasi-Machian argument. Note also the controversy about the definition of a horizon in Dirac’s cosmology.

Apart from this critical remark, one should keep in mind that Dirac’s new assumption $M \sim t^2$ does not follow directly from the LNH. It is an additional hypothesis though in the same quasi-Machian spirit. Dirac’s three possibilities of continuous creation. Either matter is created uniformly throughout space, or in special places like nuclei of galaxies and quasars, or where it exists, proportional to the amount of matter already there and of the same chemical composition. The first two possibilities would not affect our limit (17) on $G/G$, since most matter would be created in intergalactic space or in special places, and the masses in the solar system would be virtually constant. Dirac prefers the last possibility. Note that in this case it is imperative to assume that a mysterious mechanism ensures that the continuous creation reproduces the chemical composition of the matter already existing. If this were not the case, e.g. if elementary particles and not whole atoms would be created, old terrestrial rocks and meteorites would have a very high hydrogen content. E.g. rocks $3.4 \cdot 10^9$ a old should contain roughly 30 percent hydrogen. This would contradict the observations.

If we follow Dirac’s new hypothesis, the luminosity of the sun, with constant chemical composition, would change as $L \sim G^2 M^5$ and the radius of the earth’s orbit $R \sim G^{-1} M^{-1}$. This was pointed out by Teller. Then with $G \sim t^{-1}$ and $M \sim t^2$ the mean surface temperature of the earth would vary as

$$T \sim t^{1/4}.$$  \hfill (20)

Since $t_0 \lesssim 2 \cdot 10^{10}$ a, at an epoch $3.4 \cdot 10^9$ a ago we would have

$$T \lesssim 273 \text{ K}.$$  \hfill (21)

This is a temperature too low to be favourable for the origin of life at this time according to current ideas and also not compatible with geochronal data. Moreover there are severe difficulties with this selective creation of matter in connection with the crystalline structure of very ancient rocks and the spectrum of the microwave background radiation, now that its blackbody nature has been firmly established. Thus we conclude that a variation $G \sim t^{-1}$ with or without continuous creation of matter can be safely excluded.

Another theory with a varying gravitational constant is the Jordan-Brans-Dicke theory (JBD), which is also based on quasi-Machian arguments. Recent analyses of lunar laser ranging data give upper limits on a possible elongation of the moon’s orbit towards the sun due to a nongeodesical motion of planetary bodies in JBD, the Nordtvedt effect, implying $\omega > 29$ [The parameter $\omega$ is a measure of the deviation of JBD from general relativity (GR)]. In the limit $\omega \rightarrow \infty$, JBD boils down to GR. Since in JBD

$$-\dot{G}/G = \left[2/(3 + 4)\right] t_0^{-1}$$  \hfill (22)

the limit $\omega > 29$ means

$$-\dot{G}/G < 10^{-12} \text{ a}^{-1}.$$  \hfill (23)

Note that this limit comes form an analysis of periodic variations in the moon’s orbit, not from an upper limit on a secular expansion. It is theory
dependent and says that, if JBD were correct, (23) must hold. Therefore our limit (17) is not superseded by (23). In summarizing this section we can say that the prospects for a significant variation of $G$ are bleak.

III. Upper Limits on Changes of the Rest Mass of Elementary Particles

Next one might argue that the mass of elementary particles is varying with time, leaving the number of particles per comoving volume constant. If $m_e$, and therefore $m_p$ [see (7)], were larger in the past, one should expect that the energy release per nuclear reaction, the mass deficit of a heavier nucleus produced $\Delta m e^2$, were proportional to $m_p$. On the other hand the cross sections for resonant reactions are proportional to the square of the reduced de Broglie wavelength $\lambda = \hbar / m v$ (see e.g. Lang$^{22}$):

Since at constant temperature $v$ is proportional to $m^{-1/2}$ we have $\lambda \sim m^{-1/2}$, and the gain in energy release is cancelled by the decreasing cross section. Therefore we obtain the same limits for $m_p/m_p$ as for $G/G$ from fossil temperatures. This is not enough to exclude Dirac’s hypothesis, since we require only $-3 \dot{m}/m = t_0^{-1}$.

Note however that we have neglected the mass dependence of the temperature. If $m_p$ were proportional to $t^{-1}$, the sun would have had a mass of $1.2 \, M_\odot$ at $t = 3.4 \cdot 10^9$ a in the past. Detailed stellar evolution calculations are necessary to get a better upper limit on $m_p$. For the moment we conclude that Dirac’s hypothesis cannot be excluded, if elementary particle masses were variable with time. Still our limit on $\dot{m}_p/m_p$ is the only one we are aware of.

Recently Malin$^{23}$ speculated that the mass of elementary particles might change as $\dot{m}/m = -3 H_0$ or as $\dot{m}/m = 3 H_0/2$. The first possibility is clearly excluded by our limit (17). In the second case the mean surface temperature of the earth would have been around 173 K at an epoch $3.4 \cdot 10^9$ a ago, when according to current opinion life started, and only 265 K only $6.6 \cdot 10^8$ a ago when the first corals appeared. Also geochemical data give mean temperatures around 300 K in the cambrian$^{14}$. Thus a variation of $\dot{m}/m = 3 H_0/2$ can also be safely excluded.

IV. Upper Limits on the Variation of Other Quantities

We proceed now with a possible variation of the rest of the constants involved in Eq. (7) to (10) in order to check Dirac’s LNH. Gamow$^8,^{24}$ proposed that instead of $G$ the elementary charge $e$ should vary with time. Now it is clear, because of (9) that a change of $e$ alone is not compatible with the LNH. Moreover, Dyson$^{17}$ could show by an analysis of the abundances of Re$^{187}$ and Os$^{187}$ in old terrestrial rocks that the charge of the proton could only have changed by

$$|\dot{e}/e| < 2 \cdot 10^{-13} a^{-1}.$$  

If we exclude variations with cosmic time of $G$, $m_e$ and $e$, the LNH boils down to

$$\dot{c}/c = -\dot{\hbar}/\hbar$$  

and

$$3 \dot{\hbar}/\hbar = t_0^{-1}.$$  

This possibility has never been considered seriously, especially not by Dirac himself.

Finally let us discuss a few upper limits on the variation of physical constants recently published. Roberts et al. (1976)$^{25}$ and Wolfe et al. (1976)$^{26}$ derived from 21-cm and optical observations of quasars at $z = 0.5$ upper limits on the time variation of the fine structure constant $\alpha$ and the gyromagnetic factor for protons of

$$\dot{\alpha}/\alpha < 2 \cdot 10^{-12} a^{-1}$$  

and

$$(\alpha^2)'/\alpha^2 + \dot{g}_v/g_v + \dot{m}_p/m_p < 2 \cdot 10^{-14} a^{-1}.$$  

An upper limit to $\dot{\alpha}/\alpha$ has its merit in itself, but one should finally realize that it has nothing to do with Dirac’s hypothesis. In it $\alpha$ is a constant. Therefore (30) is simply an upper limit to $\dot{g}_v/g_v$, if one accepts Dirac’s hypothesis.

V. Discussion

We have shown that geochemical data on the surface temperature of the earth in the precambrian exclude Dirac’s LNH with $G \sim t^{-1}$ both with and without matter creation. Our limit on $G/G$ is one of the best obtained so far. We cannot exclude a variation of the inertial mass of elementary particles $\dot{m}/m = t_0^{-1/3}$, required as an alternative to a variation of $G$ by Dirac’s LNH. Nevertheless we can put
upper limits on $\dot{m}/m$, which has not been done before. In particular Malin's hypotheses on the variation of $\dot{m}/m$ can safely be excluded. Upper limits to changes of the elementary charge with time imply that Dirac's LNH is only tenable if either the inertial mass of elementary particles or the velocity of light and Planck's constant simultaneously are a function of time. Both possibilities were never considered by Dirac himself. On the whole we feel that it does not look bright for Dirac's LNH.

However this might be, we stress that limits on the time variation of the ratio of inertial masses of different elementary particles and of the fine structure constant have no bearing on Dirac's LNH, valuable as they are in themselves. We think that like other stimulating alternatives to Einstein's general relativity, the steady-state theory, the Jordan-Brans-Dicke theory, and Mach's principle, Dirac's large number hypothesis should be buried with appropriate honours.

6. E. Teller, Phys. Rev. 73, 801 [1948].
12. H. O. Pflug, private communication [1977].