

# The role of chaotic resonances in the Solar System

N. Murray\* & M. Holman†

\* Canadian Institute for Theoretical Astrophysics, 60 St George Street, University of Toronto, Toronto, Ontario M5S 3H8, Canada

† Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

**Our understanding of the Solar System has been revolutionized over the past decade by the finding that the orbits of the planets are inherently chaotic. In extreme cases, chaotic motions can change the relative positions of the planets around stars, and even eject a planet from a system. Moreover, the spin axis of a planet—Earth's spin axis regulates our seasons—may evolve chaotically, with adverse effects on the climates of otherwise biologically interesting planets. Some of the recently discovered extrasolar planetary systems contain multiple planets, and it is likely that some of these are chaotic as well.**

Observers have targeted solar-type stars in their searches for planets, partly motivated by the belief that these stars might harbour Earth-like planets. Although present detection techniques are unable to detect terrestrial-mass planets, they are sensitive to Jupiter-mass bodies. The original idea was that because all of these solar systems should share a common planet formation process they would result in similar collections of planets; large planets might be accompanied by smaller counterparts. However, the wide variety of orbits of the planets detected so far shows that our knowledge of planet formation is incomplete. Nevertheless, although the formation processes in other solar systems might be quite different from those that shaped our own, the long-term dynamical evolution is governed by the same principles.

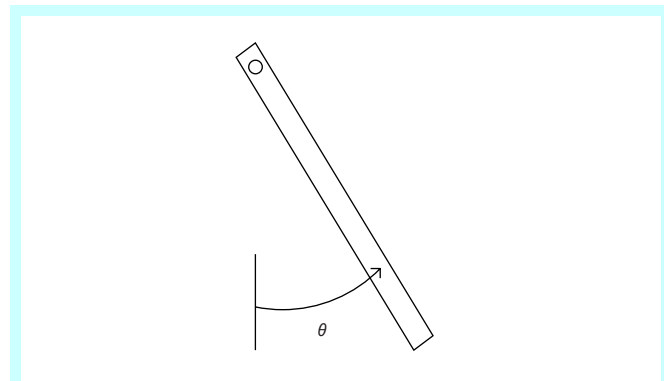
These principles have been known since Newton, but their consequences for our Solar System still surprise researchers. For nearly 300 years astronomers believed that the orbits of the planets were regular and predictable. After an initial phase of formation, the number of planets, asteroids and comets was fixed. Planets neither escaped the Solar System nor collided with each other. The discovery of chaos destroyed this traditional picture.

Our Solar System provides a plethora of examples of chaotic motion. The theory of chaos has been used to explain, and in some cases predict, the location and extent of gaps in the asteroid and Kuiper belts<sup>1–7</sup>. The theory predicts that irregularly shaped satellites such as Hyperion tumble chaotically<sup>8</sup>. The obliquity of Mars undergoes large excursions<sup>9</sup>; these are chaotic<sup>10,11</sup>. Even the obliquity of the distant future Earth may undergo chaotic evolution in 1.5–4.5 Gyr

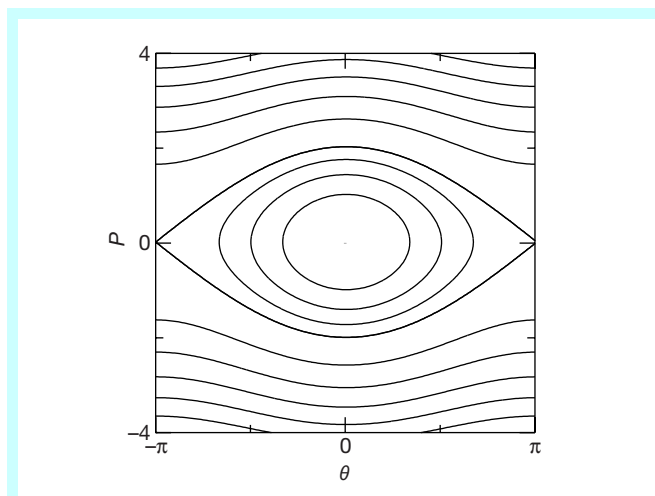
(refs 12–14). Furthermore, the chaos in the orbits of the giant planets, first observed by Sussman and Wisdom<sup>15</sup>, is due to a three-body resonance among Jupiter, Saturn and Uranus<sup>16</sup>; the locations of these planets cannot be predicted on a timescale longer than a few tens of millions of years. This supports numerical experiments that showed that our planetary system evolves chaotically<sup>15,17</sup>. The worried reader may find some comfort in that the accompanying analytic theory predicts that no planet will be ejected before the Sun dies.

## The rigid pendulum

Despite the variety and complexity of the applications mentioned above, we can introduce many of the concepts in solar system dynamics using the pendulum: phase space structure, periodic motion and stability. To describe the state of the pendulum we must specify both its position (such as the angle from the downward vertical,  $\theta$ , see Fig. 1) and velocity (such as the rate of change of the angle with time,  $\dot{\theta}$ , or the momentum  $p = m l \dot{\theta}$ , where  $m$  is the effective mass of the pendulum and  $l$  is the length). The structure of the phase space is built around fixed points (see Fig. 2). When the pendulum is hanging straight down and motionless we have a fixed point. There is another fixed point when the pendulum is straight up and motionless. Readers equipped with a pencil can immediately see that these two points are different: dangling a pencil loosely



**Figure 1** A rigid pendulum, with angle  $\theta$  measured from the vertical. The pendulum may rotate through  $360^\circ$ , unlike a simple pendulum consisting of a mass suspended on a string.



**Figure 2** The motion of a rigid pendulum traces closed curves on the phase diagram, showing the angle of pendulum,  $\theta$ , versus its angular momentum,  $p = m l \dot{\theta}$ . The stable fixed point is at  $(0,0)$ . The separatrix (see text) emanates from the unstable fixed point at  $(\pm\pi,0)$ . The shape of the separatrix resembles a cat's eye.

between two fingers is easy, but balancing a pencil upright challenges the most dextrous.

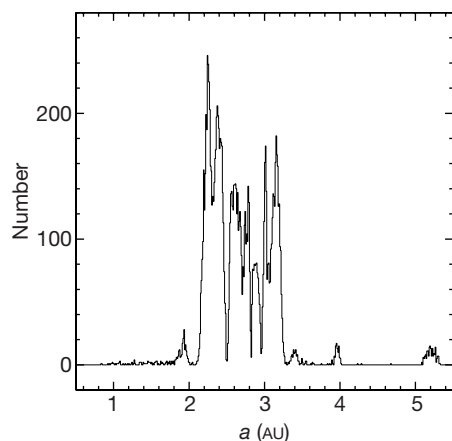
If a pendulum at the downward fixed point is nudged it begins to oscillate or librate back and forth; this fixed point is said to be stable. If a pendulum momentarily balanced on its head is gently nudged it swings down and back up on the other side; this fixed point is unstable. If the pendulum is pushed forcefully enough it will rotate periodically. Thus, we see three main regions of the phase space: circulation in one direction, libration, and circulation in the other direction. These three regions are separated by the curve that connects the unstable fixed point to itself. This curve forms the 'cat's eye' in Fig. 2, and is called the separatrix. All trajectories inside the separatrix are said to be in resonance. Trajectories well outside the 'cat's eye' have enough kinetic energy to be unaffected by the gravitational potential. They rotate with an almost fixed angular frequency. In contrast, trajectories inside the resonance are dominated by the potential; they librate. The average value of  $\dot{\theta}$  in the resonance is zero, which is a generic signature of resonant phenomena.

The separatrix is the cradle of chaos. Suppose we reach in a hand and just nudge a moving pendulum in an attempt to change the character of its motion. If the pendulum is far from the separatrix (either librating near the stable fixed point or rotating) the nudge will not change the character of the motion. However, close to the separatrix the pendulum nearly stands motionless on its head once every libration or rotation period. A nudge can then change rotation into libration, libration into rotation, or rotation in one direction into rotation in the other direction. Aside from the separatrix, all the orbits in Fig. 2 lie on closed curves. These curves are called invariant curves; similar curves appear in most dynamical systems.

In order for chaos to occur there must be at least two interacting oscillators. In the Solar System that interference is supplied by a third body, the invisible 'hand' made immanent. This interference manifests itself as zones of chaotic motion, which are the 'ghosts' of the departed separatrices. That phase space would harbour regions where trajectories are largely quasiperiodic and regions where trajectories are largely chaotic is a feature of generic dynamical systems. This characteristic is known as a divided phase space<sup>18,19</sup> (see Box 1).

### Kirkwood gaps

A series of remarkable features in the asteroid belt vividly illustrates



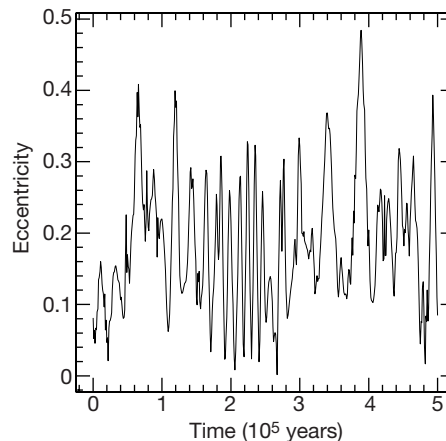
**Figure 3** The histogram of asteroids as a function of semimajor axis  $a$ . We note the distinct deficit of asteroids at, for example, 2.5, 2.82, and 2.96 AU. Gaps of this type were first noted by Kirkwood in the late 1800s. They correspond to orbital resonances with Jupiter; for example, an asteroid placed at 2.5 AU will orbit with a period equal to one-third of Jupiter's (the 3:1 resonance).

the importance of dynamical chaos in the Solar System. The distribution of semimajor axes of the orbits of the asteroids contains a number of distinct gaps (see Fig. 3). These are called Kirkwood gaps<sup>20</sup>. Kirkwood first identified them and noted that they occur at locations where the orbital period,  $T$ , which depends on the semimajor axis, would be of the form  $(p/q)T_J$ , where  $T_J$  is the orbital period of Jupiter and  $p$  and  $q$  are integers. In terms of the orbital frequencies,  $n = 2\pi/T$ ,  $qn_J - pn \approx 0$ ; in other words, this is a resonance. Nearly a century passed before it was explained<sup>2,3</sup> how a resonance produced a shortage of asteroids, or a gap. The reason for this delay can be appreciated by making a simple estimate of the strength of a resonant perturbation. The force is proportional to the mass of Jupiter, which is  $\mu_J \approx 0.001$  in units of the solar mass. Worse still, the resonant force is zero if the orbits are circular (for a planar model). If the orbits are non-circular, with an eccentricity  $e$  for the asteroid and  $e_J$  for Jupiter, the force is proportional to  $e^{p-q}e_J^s$ , where  $s$  is an integer between zero and  $|p - q|$ . For example, for a 3:1 resonance  $p = 3$  and  $q = 1$ , three force terms are possible:  $e^2$ ,  $ee_J$ , and  $e_J^2$ . Because  $e \approx e_J \approx 0.05$ , the force exerted on the asteroid by the resonance is smaller by a factor of  $5 \times 10^{-5}$  than the force exerted by the Sun, in the best possible case of a first-order resonance ( $|p - q| = 1$ ).

Laplace noted even before the discovery of the Kirkwood gaps that a resonant force adds up over many orbits, but only up to half the libration period (the time to go halfway around the 'cat's eye' in Fig. 2), which is proportional to the inverse of the square root of the resonant force. The integrated force is then larger, roughly like the square root of  $\mu_J e^{p-q}$ . This is still a small effect, so it is difficult to see how the asteroid can be ejected from the resonance.

Even if the force added up over longer times, theorems on the behaviour of dynamical systems from the 1950s and 1960s suggested that gap formation might be difficult. For example, the motion of a typical asteroid in an imaginary solar system with very small planetary masses lies on invariant curves similar to those in Fig. 2 (see Box 1). This is a specific example of the KAM theorem<sup>21-23</sup>. In exceptional cases, near the separatrix of a resonance, the theorem does not apply.

That the theorem is restricted to very small masses suggests a way to evade the dilemma; perhaps for realistic values, invariant curves at substantial distances from the separatrix are destroyed. Numerical integrations of a simplified model, incorporating only a single planet (Jupiter) on a fixed ellipse, showed that the invariant curves



**Figure 4** The orbital eccentricity of an object placed in the 3:1 resonance (at 2.5 AU) plotted as a function of time. The initial eccentricity is small, but chaotic perturbations from Jupiter force the eccentricity  $e$  of the asteroid to undergo a random walk, leading to a net increase in  $e$  and the eventual removal of the asteroid from the Solar System.

## Box 1

## Chaos and resonance overlap

Here we describe a simple model that captures many of the generic features of conservative or hamiltonian dynamical systems. As stated earlier, at least two oscillators must be interacting or coupled for chaos to occur. We can add a second oscillator to the rigid pendulum by vertically (or horizontally) moving the pivot of the pendulum periodically. Much insight can be gained by carefully examining this and similarly simple models<sup>4,6,27,28</sup>.

To describe the motion we now need two angles, say the angle of the pendulum,  $\theta$ , and the phase of the pivot,  $\psi$ . In addition we need two angle derivatives,  $\dot{\theta}$  or  $p = \dot{\theta}/\beta$ , and  $\dot{\psi}$ . The hamiltonian is

$$H = \frac{1}{2}\beta p^2 + \alpha l + [\epsilon + 2K \cos \psi] \cos \theta \quad (1)$$

$$= \frac{1}{2}\beta p^2 + \alpha l + \epsilon \cos \theta + K \cos(\theta - \psi) + K \cos(\theta + \psi)$$

The first line presents an amplitude-modulated pendulum; the second presents three phase-modulated pendulums. See ref. 25 for a detailed development. Several constants depend on the physical characteristics of the driven pendulum:  $\beta$  is related to the moment of inertia of the pendulum;  $2K$  is the amplitude of the driving; and  $\epsilon$  is proportional to the gravitational acceleration. The constant  $\alpha$  is the driving frequency or  $\dot{\psi}$ , and  $l$  is the negative of the total energy of the pendulum.

There are three resonances (cosine terms) which dominate the motion when their arguments are slowly varying:  $\dot{\theta} \approx 0$ ,  $\dot{\theta} \approx \dot{\psi}$ , and  $\dot{\theta} \approx -\dot{\psi}$ . We can visualize the trajectories by plotting  $p$  versus  $\theta$  whenever the driving reaches a specific phase,  $\psi = 0$  for example. The resulting diagram is called a surface of section.

The figure shows surfaces of section for  $\beta = 1$ ,  $\epsilon = -1$ , and  $K = -0.3$ . Panels **a**, **b** and **c** correspond to decreasing driving frequencies of  $\alpha = 7$ , 3 and 0.3, respectively. Here, the natural oscillation frequency of the pendulum is  $\sqrt{|\epsilon|/\beta} = 1$ . The resonant islands are located at  $p = 0$  and  $\beta p = \pm \dot{\psi} = \pm \alpha$ . The centre resonant island has a half-width of  $\Delta P = 2\sqrt{|\epsilon|/|\beta|} = 2$  and the other two have half-widths of  $\Delta P = 2\sqrt{K/\beta} \approx 1.1$ . So, for the highest driving frequency (**a**) the centres of the resonant 'islands' are separated by more than their widths. Aside from the additional resonant islands, two new features appear. First, two chains of secondary islands can be seen. Second, the separatrix of the main island has broadened into a fuzzy zone. This zone was traced out by a single trajectory. Each of the separatrices shows such a chaotic zone, although they are too small to discern. Nevertheless, what appear to be invariant curves corresponding to regular trajectories can still be seen. That both regular and chaotic trajectories should appear on the same surface of section is a generic feature of a divided phase space<sup>16,19,28</sup>.

In **b** the driving frequency has been reduced,  $\alpha = 3$ , such that now the islands, ignoring their interaction, just touch each other. As the driving frequency approaches the natural oscillation frequency of the pendulum, the small chaotic zone seen in **a** envelopes the three resonant islands. This is the result of the beginning of resonance overlap<sup>28</sup>.

In **c** the driving frequency has been further reduced,  $\alpha = 0.3$ , such

that now the resonances have widths much larger than their separations. The motion is chaotic near the separatrix, which appears as a single thick band surrounding a stable island, and is surrounded by invariant curves. As can be seen from the hamiltonian, the three resonant terms can also be viewed as a single resonance with an oscillating width,  $\Delta P = 2\sqrt{|\epsilon| + 2K \cos \psi}/\beta$ . The width ranges from  $\Delta P \approx 1.3$  to  $\Delta P \approx 2.5$ , for the chosen values of  $\epsilon$  and  $K$ . Here, the chaotic zone is roughly that region over which the separatrix sweeps. This is termed modulational chaos<sup>51</sup> or adiabatic chaos<sup>4</sup>.

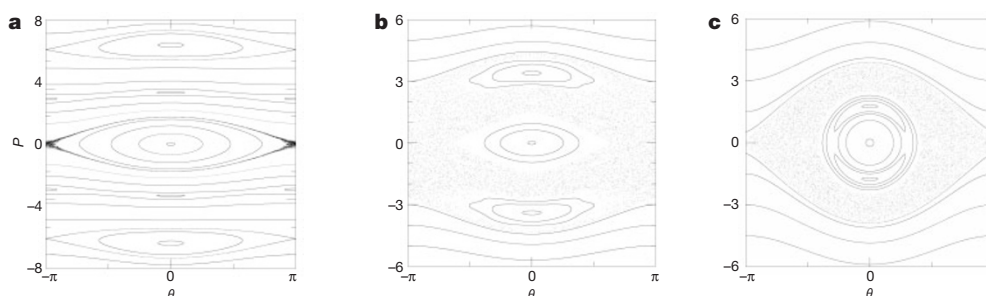
In discussing the origin of chaotic motion in the main text, we have treated the region of overlap as though it had no structure. This simplification is justified in some cases (for example in high-order resonances in the asteroid belt) and for some applications (for calculating Lyapunov times), but is not always adequate.

For example, the eccentricity of asteroids in the 3:1 resonance could librate around two different centres, one at the classical 'forced eccentricity'  $e_f \approx 0.05$  and a second at  $e \approx 0.15$  (ref. 3). The chaotic region surrounded the separatrix between these two points, allowing asteroids to pass from low ( $\sim 0.05$ ) to high ( $\sim 0.3$ ) eccentricities, where they could be removed by encounters with Mars. Later a third libration centre at even higher eccentricity was discovered, connected to the first two by chaotic orbits<sup>52</sup>. This allowed asteroids to reach  $e > 0.06$ ; such objects are removed from the resonance, typically by plunging into the Sun. This motion is made possible by the resonance overlap, but describing it naturally requires knowledge of the underlying phase space structure; or we can perform numerical simulations.

Direct brute-force numerical integration indicates that plunging into the Sun is the most likely fate of bodies injected into the 3:1 resonance, with the remainder being pushed beyond Saturn<sup>53</sup>. These integrations find a removal time of about one million years. There is also direct observational evidence that this interpretation is correct because the boundaries of the chaotic zone closely match the boundaries to observed distribution of asteroids<sup>3</sup>. Similar statements can be made about the other Kirkwood gaps and gaps in the outer asteroid belt<sup>6</sup> (see Fig. 6).

A more realistic model for the 3:1 mean motion resonance, still for a single perturbing planet, would include terms of order  $I^2$ , and allow for (different)  $I$  dependences in the factors  $K$  multiplying the cosine terms involving  $\psi$ . Such a model reproduces the two libration centres in the  $\psi$ ,  $I$  motion found by Wisdom. Because the analytic model that we, and Wisdom, employ is of low order in  $e$ , it misses the third libration centre found in ref. 52.

More complete dynamical models, in particular those that include the effect of Saturn on Jupiter's orbit, show that secular resonances, that is, resonances between the precession periods of the asteroid and those of Jupiter and Saturn, overlap inside low-order mean motion resonances such as the 2:1<sup>54,55</sup>. This overlap can enhance the rate of diffusion. Even this is not the entire story; Jupiter and Saturn are near (but not in) a 5:2 resonance. This 'great inequality' enhances the rate of diffusion of asteroids in the 2:1 mean motion resonance with Jupiter<sup>56</sup>.



**Box 1 Figure** Surface of section for the driven pendulum. **a**, Driving frequency much faster than the natural pendulum oscillation frequency. **b**, Driving frequency approaching the natural pendulum

oscillation frequency. **c**, Driving frequency less than the natural pendulum oscillation frequency.

were destroyed in the vicinity of the 2:1 resonance<sup>1</sup>. This removed the barrier of the KAM theorem. Unfortunately, these  $\sim 100,000$  yr integrations<sup>1</sup>, although they established the presence of chaotic orbits, did not give any hint as to how the irregular motion led to the removal of an asteroid from the resonance<sup>1,24</sup>.

Then Wisdom showed by direct numerical integration that the eccentricity of small bodies placed in the 3:1 gap alternates chaotically between periods of low and moderate eccentricity as the result of perturbations from Jupiter<sup>2</sup> (see Fig. 4). This result relied upon the development of an algebraic mapping that could efficiently follow the long-term motion of asteroids in the 3:1 mean motion resonance with Jupiter for a million years or longer. This is a recurring theme in Solar System dynamics: more efficient algorithms and faster computers permit longer, more accurate integrations that often reveal unexpected dynamical results.

What is the source of the chaos in the 2:1 and 3:1 mean motion resonances? It has long been known that each mean motion resonance is composed of several individual resonant terms<sup>25</sup>. For example in the 3:1 resonance there are three such terms in the planar case, as mentioned above. All of these resonances are weak, but they are also close together<sup>2-4,26</sup>. It is the close proximity of multiple, albeit weak, resonances which invalidates the simple estimate for the integrated force given above. A single resonance produces a small force integrated over a libration period, and a simple 'cat's eye' in phase space with a smooth separatrix. Adding a second weak resonance breaks up this smooth separatrix in a chaotic region (see Box 1). The separatrix is broken because the second resonance 'nudges' the pendulum represented by the first resonance, altering the motion, particularly that of orbits that pass near the unstable fixed point<sup>27</sup>. As a result the motion from one passage near the fixed point to the next is uncorrelated. The resonant force is still small compared to the force exerted by the Sun, and produces only small changes over a libration period. But over times longer than a libration period the asteroid experiences uncorrelated forces.

As a result of these uncorrelated forces the motion in the chaotic zone differs from that in a regular region; the chaotic motion is very sensitive to initial conditions. Suppose we take two copies of our interacting oscillators, released from slightly different positions. In one case we start near a stable fixed point, well away from the separatrix. Here the motion is regular and the two copies of the apparatus separate from each other linearly (or polynomially) in time (see Fig. 5). In the other case we start the two copies near an unstable fixed point. Now the separation increases exponentially in time. The timescale for this separation is called the Lyapunov time,  $T_L$ .

We can estimate the Lyapunov time of an asteroid placed artificially in the 3:1 mean motion resonance. There are only two relevant timescales in the problem: the libration period and the precession period. From our argument above, large-scale chaos occurs when the separation of the 'islands' (or 'cat's eyes') is comparable to their widths<sup>28</sup>, that is, when the resonances overlap. Another way to say this is that large-scale chaos occurs when the libration and precession periods of a resonant asteroid are similar. Then the Lyapunov time must be equal to the libration period—the only timescale in the problem<sup>29</sup>. In the case of the 3:1 resonance the precession period is  $3 \times 10^4$  years, whereas the numerically determined Lyapunov time is  $1.4 \times 10^4$  years.

The two types of trajectories differ in more than their rates of separation. The chaotic trajectory can explore a larger volume of phase space than is accessible by a regular trajectory. For asteroids this extra measure of freedom can be dangerous, opening routes by which they can be ejected or cannibalized by planets<sup>2-5</sup> (it's world-eat-world out there).

In higher-order resonances, where the phase-space structure is relatively simple, the uncorrelated or chaotic forces lead to simple random walks involving only small exchanges of energy between the planet and the asteroid, because the semimajor axes of both bodies

are nearly fixed by the resonance condition. This is not so for the angular momentum; the torques associated with the uncorrelated forces produce a random walk in the angular momentum of both bodies. This leads to a classic gambler's ruin problem with the asteroid playing the part of the gambler and the planet taking the role of the house. If either body loses most of its angular momentum, that body's orbit becomes highly eccentric and subject to collisions with other objects in the Solar System or ejection—which is analogous to bankruptcy. The planet, with its much larger mass, has more capital (angular momentum) than the asteroid. Even worse, the asteroid is like a gambler forced to limit his winnings to what he can carry in his pockets; because of its small mass the asteroid cannot absorb enough angular momentum to produce a substantial change in the orbit of the planet. The two trade angular momentum back and forth, but when the inevitable happens and the asteroid loses its small stock of angular momentum, it is abruptly escorted from the cosmic casino.

We can estimate the time needed to remove the asteroid. Because the motion is chaotic we can treat the resonance angle as a random variable, which drives random changes in the eccentricity. In the case of the 3:1 resonance the changes in  $e^2$  are of order  $\Delta e^2 \approx (T_L/T_J)(M_J/M_\odot)e^2$ , using the Lyapunov time,  $T_L$ , as the interval associated with the random changes ( $M_J$ ,  $T_J$ , and  $M_\odot$  are Jupiter's mass, Jupiter's orbital period, and the Sun's mass, respectively). The short-term average eccentricity diffuses to large values in a time of about one million years.

In some cases, such as the asteroid Helga in the 12:7 resonance, the motion is chaotic with a very short Lyapunov time, but the diffusion time is comparable to or larger than the age of the Solar System. When applied to Helga this theory predicts that the asteroid should survive for about 8 billion years. Direct numerical integrations agree with this prediction<sup>6</sup>.

We note that the largest gap, that between the 1:1 resonance harbouring the Trojan asteroids, and the Hildas in the 3:2 resonance, is the result of the overlap of distinct mean motion resonances<sup>30</sup>; for example the 4:3 resonance overlaps with the 5:4, the 5:4 with the 6:5, and so forth.

It is believed that most meteorites come from one of two sources, the 3:1 mean motion resonance and the  $\nu_6$  secular resonance<sup>4,31,32</sup>. The latter involves a resonance between the precession frequency of the apsidal line of the asteroid and the sixth fundamental secular frequency of the Solar System (which is very roughly Saturn's precession frequency). Roughly equal numbers of meteorites come from each type of resonance. Until recently there was one difficulty with this idea. The cosmic ray exposure ages (a measure of the delivery time) of the stony meteorites are typically 20 Myr (ref. 33), whereas the delivery time from the 3:1 resonance is much shorter, 1 Myr. However, recent simulations<sup>34</sup> suggest a way out of this dilemma: most meteorites, which are produced by collisions between larger bodies in the asteroid belt, are not injected directly into either the 3:1 or the  $\nu_6$  resonance. Rather, they are placed in the vicinity of the resonance, and then dragged into the unstable region by the Yarkovsky effect<sup>34</sup>, which arises due to the anisotropic thermal radiation from those regions of the fragment which are exposed to the Sun. With the addition of this slow dynamical precursor, the theory of meteorite delivery appears to be complete.

### Chaotic spin-orbit resonances

One of the most important examples of chaos is afforded by the evolution of planetary spins. This chaos is produced by resonances between spin and orbit precession periods. In such resonances, the asphericity of the planet couples to the non-axisymmetric perturbation produced by orbital eccentricity or inclination. The small deviations from perfect sphericity, typically a few parts in a thousand for larger bodies like planets, leads to significant exchanges of energy and angular momentum between orbital and



rotational (or spin) motions of satellites and even planets. Resonances between spin period and orbital period are common in the Solar System, the Moon being a prominent example. The phase space occupied by such resonances is small, making it unlikely that so many bodies formed in resonance. This paradox is explained by the fact that dissipative effects tend to drive bodies into these resonances, where the motion is stabilized<sup>35,36</sup>. As a result, most resonant satellites are currently deep in their respective one-to-one spin-orbit resonances, where the motion is regular; however, in one case, Saturn's satellite Hyperion, the chaotic motion may have persisted until the present<sup>8</sup>.

Because solar tides are so weak, dissipative effects tend to be less important for planetary spins, allowing for richer dynamics. Ward<sup>37</sup> showed that the angle between the spin and orbital axes of Mars (the obliquity) varies by  $\pm 13.6^\circ$  around its average of  $24^\circ$  over millions of years. These variations are a result of a resonance between the precession of the spin axis and a combination of orbital precession frequencies; improvements in orbital models resulted in an increased variation of  $\pm 20^\circ$  (ref. 9). Numerical integrations then showed that the obliquity of Mars is evolving chaotically, and varies over an even larger range<sup>10,11</sup>. Such a variation has profound, but as yet poorly understood, implications for climate variation. It is likely that the spin axes of Mercury and Venus underwent chaotic variations in the past<sup>11</sup>.

The tilt of the Earth, currently  $23^\circ$ , will also increase in the future<sup>12</sup>; the Moon will evolve outward under the influence of the tides, resulting in a decrease in the precession rate of the Earth. Eventually the precession rate will become resonant with yet another combination of orbital precession rates. Once again, numerical integrations show that the Earth's obliquity will vary chaotically<sup>13,14</sup>. The tilt of the Earth's axis may increase to  $90^\circ$ . The effect on our climate is hard to estimate, but the result is unlikely to be pleasant.

### Three-body resonances

Numerical integrations of main-belt asteroids show that a substantial fraction of these bodies have chaotic orbits with rather short Lyapunov times ( $\sim 10^5$  years)<sup>6,38,39</sup>. Most of this chaotic motion is not associated with any two-body resonance. Instead it is the result of the interaction among three bodies: the asteroid, Jupiter, and either Mars in the inner asteroid belt or Saturn in the main or outer asteroid belt<sup>38,39</sup>. As with Helga, the Lyapunov times can be quite short, but the diffusion times are comparable to or longer than the age of the Solar System.

Three-body resonances arise when one planet, such as Jupiter, is perturbed by a second, such as Saturn. The orbit of Jupiter is then no longer a simple keplerian ellipse. The potential experienced by the asteroid in Jupiter's gravitational field is given by an expression formally equivalent to the two-body case, but Jupiter's orbital elements now vary with time. This variation introduces a whole new suite of frequencies into the potential experienced by the asteroid; in addition to all the harmonics of Jupiter's period, all the harmonics of Saturn's orbital period appear as well. Three-body resonances have also been considered for the galilean and uranian satellites, as well as in ring systems<sup>40–42</sup>.

These new terms in the potential have much smaller amplitudes than do two-body resonant terms that have the same number of powers of eccentricity. Three-body resonances are proportional to the product of the masses of the two perturbing bodies,  $M_J$  and  $M_S$  in our example, rather than just  $M_J$  as in a two-body resonance.

The structure of a three-body resonance is similar to that of a two-body resonance—multiple, very narrow components separated in semimajor axis by an amount that depends on the precession frequencies of the bodies involved. When the separation is comparable to or smaller than the width of the individual resonances, the motion in the immediate vicinity of the resonance will be chaotic. We can then estimate the Lyapunov time and the

diffusion time in the same manner as for two-body resonances.

We have searched the catalogue of asteroids for evidence of gaps at the location of a number of the stronger three-body resonances. We have not been able to identify a candidate gap.

We can check the theory in other ways. For example, some asteroids on the inner edge of the belt (near Mars) could have formed in three-body resonances involving the asteroid, Mars and Jupiter. If the diffusion time of some of these asteroids is comparable to the age of the Solar System, we should be able to find objects that are about to be removed from the belt owing to close encounters, or possibly collisions, with either Mars or Earth. Such bodies may have already been discovered: the so-called near-Earth asteroids<sup>43</sup>. These bodies have been a puzzle for astronomers, as in their present orbits they have very short lifetimes, typically millions of years. Because the Solar System is billions of years old, objects with lifetimes of millions of years should have vanished long ago, unless there is some way to replenish the supply. The theory of three-body resonances offers a possible mechanism.

### Chaos among the giant planets

Even trajectories outside but near a resonance can be affected by its presence. In the century following Newton's publication of his law of universal gravitation, astronomers noted that the positions of Jupiter and Saturn deviated from their predicted positions by some 30 minutes of arc. The difference became known as the great inequality. Laplace noted that the predictions did not take into account the influence of Saturn on Jupiter's orbit, and vice versa. On the face of it this seemed reasonable, because the mass of either planet was less than one one-thousandth that of the Sun. Early astronomers had made the naive estimate outlined above in our discussion of the Kirkwood gaps, which integrates this tiny force over the interval between successive conjunctions of the planets, a time somewhat longer than  $T_J$ . But Laplace realized that the orbital period of Saturn was almost exactly  $5/2$  times that of Jupiter ( $2/T_J - 5/T_S \approx 1/85T_J$ ). The perturbations accumulate over a much longer interval ( $85T_J$ ), permitting a larger exchange of energy and angular momentum between the two planets. The change in the predicted position of the planet on the sky was roughly  $85 \times 85$  times larger than the simple estimate would indicate. With this result, Laplace was able to reconcile the observations with the prediction of the law of gravity.

This discovery strongly affected Laplace's views regarding determinism, reflected in his well known statement<sup>44</sup>:

The present state of the system of nature is evidently a consequence of what it was in the preceding moment, and if we conceive of an intelligence that at a given instant comprehends all the relations of the entities of this universe, it could state the respective position, motions, and general affects of all these entities at any time in the past or future.

Physical astronomy, the branch of knowledge that does the greatest honour to the human mind, gives us an idea, albeit imperfect, of what such an intelligence would be.

This view of the world passed into common currency: the clockwork motion of the planets became the epitome of regularity.

This view is wrong. It is wrong because it ignores the delicate nature of the separatrix, the cradle of chaos. The giant planets provide further evidence of the power of three-body resonances, the subtlety of natural phenomena, and the difficulty of interpreting the fruits of scientific enquiry.

The *tour de force* numerical integrations of the outer planets in 1988 (ref. 45) shattered the clockwork. Using the 'Digital Orrery', a parallel computer built specifically for the task and now part of the Smithsonian collection in Washington DC, they followed the motion of the four giant planets and Pluto (as a test particle) over 845 Myr. To the surprise of all at the time, Pluto's orbit was chaotic with  $T_L \approx 10$  million years.

The breakthrough was followed by a re-examination of the orbital evolution of the full Solar System. Longer integrations<sup>46,47</sup> of an approximate model of the Solar System (the orbits were averaged, and Pluto was ignored) showed that the Solar System itself was chaotic. Laskar suggested that secular resonances involving the terrestrial planets were responsible for the chaotic motion<sup>46,47</sup>. Subsequent integrations of a complete model<sup>15</sup> confirmed that the full Solar System was chaotic with  $T_L \approx 5$  Myr. Sussman and Wisdom also found that the four giant planets by themselves appeared to form a chaotic system<sup>15</sup>.

The chaos amongst the four giant planets arises from a three-body resonance involving Jupiter, Saturn and Uranus<sup>16</sup>. The orbital period of Uranus is nearly seven times that of Jupiter. In terms of the orbital frequency  $n = 2\pi/T$ , the difference  $n_J - 7n_U$  is equal to  $5n_S - 2n_J$ . The relevant resonant terms in the potential experienced by Uranus are proportional to  $M_J M_S / (2 - 5n_S/n_J)$ . This ‘small denominator’,  $(2 - 5n_S/n_J) \approx 1/85$ , enhances the potential experienced by Uranus by a factor of 85. It was of course this small denominator that Laplace used to explain the origin of the great inequality, and which is implicated in the removal of objects from the 2:1 resonance in the asteroid belt.

The width of a single component of this three-body resonance is tiny, comparable to the radius of Uranus, but so is the separation between components; the resonances just overlap. The predicted Lyapunov time is  $T_L \approx 10$  million years. Because Uranus is so much less massive than either Jupiter or Saturn, its orbital angular momentum is substantially less than that of the two gas giants. By the analogy above, it is hence subject to the gambler’s ruin. The diffusion time (roughly the time before Uranus is ejected) is  $10^{18}$  years, much longer than the current age of the Universe. The theory also predicts the location and Lyapunov times of other chaotic zones near the present orbit of Uranus. Detailed numerical integrations verify the presence and Lyapunov times of all these zones, as well as illustrating the transitions between libration and rotation in the relevant resonances. Current computational resources are inadequate to test the predicted diffusion time.

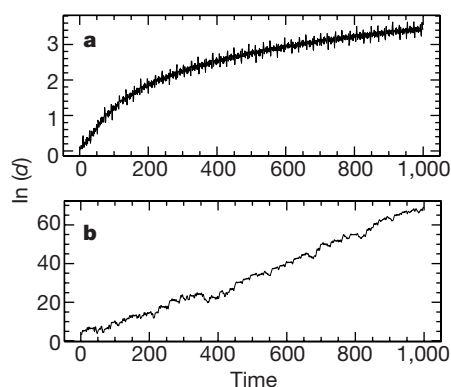
### Chaos among the terrestrial planets

As noted above, numerical integrations of the full Solar System, inner planets included, show evidence of chaos with a Lyapunov time of  $T_L \approx 5$  million years<sup>15,17</sup>. Unlike the outer planets, the source of this chaos has not been convincingly established. Laskar pointed to what are called “secular resonances” between the terrestrial planets as a candidate source of the chaos<sup>46</sup>. He

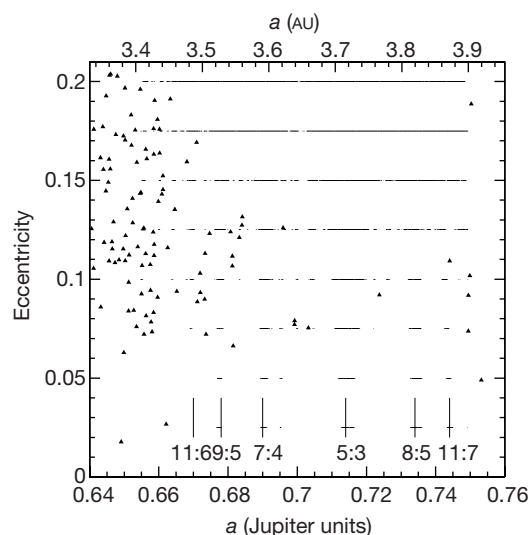
found what appears to be an alternation between circulation and libration in the angles  $\sigma_1 \equiv (\bar{\omega}_1 - \bar{\omega}_5) - (\Omega_1 - \Omega_2)$  and  $\sigma_2 \equiv 2(\bar{\omega}_4 - \bar{\omega}_3) - (\Omega_4 - \Omega_3)$ . In these relations  $\bar{\omega}_1$  refers to the orientation of Mercury’s apsidal line, which is the line from the Sun to the point of Mercury’s orbit closest to the Sun, whereas  $\Omega_1$  refers to the orientation of Mercury’s nodal line; the latter is defined by the intersection of the orbital plane of Mercury with the orbital plane of the Earth. Similar definitions apply to the elements for Venus (subscript 2), Earth (3), Mars (4), and Jupiter (5). However, these two resonances do not interact directly with each other, so by themselves they are unlikely to produce large-scale chaos. Later integrations<sup>47</sup> identified a third resonance,  $\sigma_3 \equiv (\bar{\omega}_4 - \bar{\omega}_3) - (\Omega_4 - \Omega_3)$ . Laskar found that it librated while  $\sigma_2$  rotated, and vice versa. As  $\sigma_2$  and  $\sigma_3$  involve the same degrees of freedom, they are a more promising candidate for overlapping resonances. Sussman and Wisdom<sup>15</sup>, employing a more realistic (unaveraged) model, confirmed the alternate libration and rotation of  $\sigma_2$ , but never saw  $\sigma_3$  librate; this does not rule out the possibility that the overlap of the associated resonances causes the chaos. However, their assessment that “no dynamical mechanism for the observed chaotic behaviour of the Solar System has been clearly demonstrated” seems warranted, at least for the terrestrial planets. Without a clear identification of the source of the chaos it is not possible to use an analytic development, such as was used for the outer planets, to confirm the Lyapunov time and then estimate the timescale for diffusion of the system.

### Chaos in other planetary systems

By analogy with our Solar System, multiplanet systems seem likely. In fact three Jupiter-mass objects orbit Upsilon Andromeda<sup>48</sup>. Any multiple-planet system is subject to the instabilities described in this Review. All of the known planetary systems orbiting solar-type stars have ages of  $10^9$  years or greater. This, together with the copernican assumption that we are not observing the system at a privileged time, such as immediately after a recent planetary ejection, can be used to put limits on the masses and orbital elements of the planets, quantities which cannot be tightly constrained or even directly observed using current techniques<sup>49,50</sup>. Dynamical constraints can



**Figure 5** The distance between two initially nearly identical initial conditions for two interacting nonlinear oscillators. **a**, The motion is near a stable fixed point; **b**, the motion starts near an unstable fixed point. In **b** (but not in **a**) the two initial conditions separate exponentially with time, and are said to be chaotic.



**Figure 6** The location of asteroids in the outer asteroid belt in the (eccentricity, semimajor axis) plane. Lyapunov times have been computed as a function of the semimajor axis for several values of eccentricity. Points are plotted when the Lyapunov time is below a threshold value. These indicate the chaotic zones. It is clear that the known asteroids avoid these regions.

also be placed on the existence of smaller (Earth-mass) bodies in orbits near those of the Jupiter-mass objects. Observational searches for such low-mass companions will have to await improved techniques. □

1. Giffen, R. A study of commensurable motion in the asteroid belt. *Astron. Astrophys* **23**, 387–403 (1973).
2. Wisdom, J. The origin of the Kirkwood gaps: a mapping for asteroidal motion near the 3/1 commensurability. *Astron. J.* **87**, 577–593 (1982).
3. Wisdom, J. Chaotic behavior and the origin of the 3/1 Kirkwood gap. *Icarus* **56**, 51–74 (1983).
4. Wisdom, J. A perturbative treatment of motion near the 3/1 commensurability. *Icarus* **63**, 272–289 (1985).
5. Wisdom, J. Meteorites may follow a chaotic route to earth. *Nature* **315**, 731–733 (1985).
6. Holman, M. & Murray, N. Chaos in high-order mean motion resonances in the outer asteroid belt. *Astron. J.* **112**, 1278–1293 (1996).
7. Malhotra, R., Duncan, M. J. & Levison, H. F. in *Protostars and Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S. S.) 1231 (Univ. Arizona Press, Tucson, 2000).
8. Wisdom, J., Peale, S. J. & Mignard, F. The chaotic rotation of Hyperion. *Icarus* **58**, 137–152 (1984).
9. Ward, W. R. & Rudy, D. J. Resonant obliquity of Mars? *Icarus* **94**, 160–164 (1991).
10. Touma, J. & Wisdom, J. The chaotic obliquity of Mars. *Science* **259**, 1294–1297 (1993).
11. Laskar, J. & Robutel, P. The chaotic obliquity of the planets. *Nature* **361**, 608–612 (1993).
12. Ward, W. R. Comments on the long-term stability of the earth's obliquity. *Icarus* **50**, 444–448 (1982).
13. Laskar, J., Joutel, F. & Robutel, P. Stabilization of the earth's obliquity by the moon. *Nature* **361**, 615–617 (1993).
14. Neron de Surgy, O. & Laskar, J. On the long term evolution of the spin of the Earth. *Astron. Astrophys.* **318**, 975–989 (1997).
15. Sussman, G. J. & Wisdom, J. Chaotic evolution of the solar system. *Science* **257**, 56–62 (1992).
16. Murray, N. & Holman, M. The origin of chaos in the outer solar system. *Science* **283**, 1877–1881 (1999).
17. Laskar, J. A numerical experiment on the chaotic behaviour of the solar system. *Nature* **338**, 237–238 (1989).
18. Chirikov, B. V. Resonance processes in magnetic traps. *Plasma Phys.* **1**, 253–260 (1960).
19. Hénon, M. & Heiles, C. The applicability of the third integral of motion: Some numerical experiments. *Astron. J.* **69**, 73–79 (1964).
20. Kirkwood, D. *Meteoritic Astronomy: A Treatise on Shooting-stars, Fireballs, and Aerolites* 105–111 (Lippincott, Philadelphia, 1867).
21. Kolmogorov, A. N. On conservation of conditionally periodic motions under small perturbations of the Hamiltonian. *Dokl. Akad. Nauk* **98**, 527–530 (1953).
22. Arnold, V. I. Proof of A. N. Kolmogorov's theorem on the preservation of quasi-periodic motions under small perturbations of the Hamiltonian. *Uspehi Mat. Nauk.* **18**, 13–40 (1963); English translation *Russian Math. Surv.* **18**, 9–36 (1963).
23. Moser, J. On invariant curves of area-preserving mappings of an annulus. *Nachr. Akad. Wiss. Göttingen Math. Phys. K* **1**, 87–120 (1962).
24. Froeschle, C. & Scholl, H. The stochasticity of peculiar orbits in the 2/1 Kirkwood gap. *Astron. Astrophys.* **93**, 62–66 (1981).
25. Peirce, B. Development of the perturbative function of planetary motion. *Astron. J.* **1**, 1–8 (1849).
26. Goldreich, P. in *Cosmology and Astrophysics* (eds Terzian, Y. & Bilson, E. M.) 121–129 (Cornell Univ. Press, Ithaca, 1982).
27. Lichtenberg, A. J. & Lieberman, M. A. *Regular and Chaotic Dynamics* 245–264 (Springer, New York, 1992).
28. Chirikov, B. V. A universal instability of many-dimensional oscillator systems. *Phys. Rep.* **52**, 263–379 (1979).
29. Murray, N. & Holman, M. Diffusive chaos in the outer asteroid belt. *Astron. J.* **114**, 1246–1259 (1997).
30. Wisdom, J. The resonance overlap criterion and the onset of stochastic behavior in the restricted three-body. *Astron. J.* **85**, 1122–1133 (1980).
31. Wetherill, G. W. Where do the Apollo objects come from? *Icarus* **76**, 1–18 (1988).
32. Morbidelli, A. & Gladman, B. Orbital and temporal distributions of meteorites originating in the asteroid belt. *Meteor. Planet. Sci.* **33**, 999–1016 (1998).
33. Marti, K. & Graf, Cosmic-ray exposure history of ordinary chondrites. *Annu. Rev. Earth Planet. Sci.* **20**, 221–243 (1992).
34. Vokrouhlický, D. & Farinella, P. Efficient delivery of meteorites to the Earth from a wide range of asteroid parent bodies. *Nature* **407**, 606–608 (2000).
35. Darwin, G. H. On the precession of a viscous spheroid and on the remote history of the Earth. *Phil. Trans. R. Soc. Lond.* **170**, 447–538 (1879).
36. Goldreich, P. Final spin states of planets and satellites. *Astron. J.* **71**, 1–7 (1966).
37. Ward, W. R. Large-scale variations in the obliquity of Mars. *Science* **181**, 260–262 (1973).
38. Murray, N., Holman, M. & Potter, M. On the origin of chaos in the asteroid belt. *Astron. J.* **116**, 2583–2589 (1998).
39. Nesvorný, D. & Morbidelli, A. Three-body mean motion resonances and the chaotic structure of the asteroid belt. *Astron. J.* **116**, 3029–3037 (1998).
40. Tisserand, F. *Traité de Mécanique Céleste* Tome IV (Gauthier-Villars, Paris, 1896).
41. Goldreich, P. & Nicholson, P. Revenge of tiny Miranda. *Nature* **269**, 783–785 (1977).
42. Aksnes, K. in *Planetary Rings* (ed. Brahic, A.) 479–487 (Proceedings of IAU Colloq. 75, Cepadues, Toulouse, 1984).
43. Morbidelli, A. & Nesvorný, D. Numerous weak resonances drive asteroids toward terrestrial planet's orbits. *Icarus* **139**, 295–308 (1999).
44. Laplace, P. S. *Oeuvres Complètes de Laplace* Vol. 8, 144–145; (Imprimerie Royale, Paris, 1843–47); translation Gillispie, C. C., Fox, R. & Grattan-Guinness, I. *Pierre-Simon Laplace* (Princeton Univ. Press, 1997).
45. Sussman, G. J. & Wisdom, J. Numerical evidence that the motion of Pluto is chaotic. *Science* **241**, 433–437 (1988).
46. Laskar, J. The chaotic motion of the solar system—A numerical estimate of the size of the chaotic zones. *Icarus* **88**, 266–291 (1990).
47. Laskar, J. in *Proc. IAU Symposium 152* (ed. Ferraz-Mello, S.) 1–16 (1990).
48. Butler, R. P. *et al.* Evidence for multiple companions to  $\nu$  Andromedae. *Astrophys. J.* **526**, 916–927 (1999).
49. Laughlin, G. & Adams, F. C. Stability and chaos in the  $\nu$  Andromedae planetary system. *Astrophys. J.* **526**, 881–889 (1999).
50. Rivera, E. J. & Lissauer, J. J. Stability analysis of the planetary system orbiting  $\nu$  Andromedae. *Astrophys. J.* **530**, 454–463 (2000).
51. Tennyson, J. in *Nonlinear Dynamics in Beam-Beam Interactions* (eds Month, M. & Herrera, J. C.) 158–193 (AIP Conf. Proc. 57, AIP, New York, 1979).
52. Ferraz-Mello, S. & Klafke, J. C. in *Predictability, Stability, and Chaos in N-body Dynamical Systems* (ed. Roy, A. E.) 177–184 (Plenum, New York, 1991).
53. Gladman, B. J. *et al.* Dynamical lifetimes of objects injected into asteroid belt resonances. *Science* **277**, 197–201 (1997).
54. Henrard, J. & Lemaitre, A. A perturbation method for problems with two critical arguments. *Celest. Mech.* **39**, 213–238 (1986).
55. Morbidelli, A. & Moons, M. Secular resonances in mean motion commensurabilities: the 2:1 and 3:2 cases. *Icarus* **102**, 316–332 (1993).
56. Ferraz-Mello, S., Michtchenko, T. A. & Roig, F. The determinant role of Jupiter's great inequality in the depletion of the Hecuba gap. *Astron. J.* **116**, 1491–1500 (1998).

## Acknowledgements

We thank P. Goldreich for helpful conversations. This research was supported by NSERC of Canada and NASA.

Correspondence and requests for materials should be addressed to M.H. (e-mail: mholman@cfa.harvard.edu).